Broadband terahertz photonics

by

David Patrick Burghoff

B.S., University of Illinois at Urbana-Champaign (2007)
S.M., Massachusetts Institute of Technology (2009)

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

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

Author .....
Department of Electrical Engineering and Computer Science
August 29, 2014

Certified by .....
Qing Hu
Professor
Thesis Supervisor

Accepted by .....
Professor Leslie A. Kolodziejski
Chair, Department Committee on Graduate Students

Broadband terahertz photonics

by

David Patrick Burghoff

Submitted to the Department of Electrical Engineering and Computer Science
on August 29, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

In recent years, quantum cascade lasers have emerged as mature semiconductor sources of light in the terahertz range, the frequency range spanning 1 to 10 THz. Though technological development has pushed their operating temperatures up to 200 Kelvin and their power levels up to Watt-level, they have remained unsuitable for many applications as a result of their narrow spectral coverage. In particular, spectroscopic and tomographic applications require sources that are both powerful and broadband. Having said that, there is no fundamental reason why quantum cascade lasers should be restricted to narrowband outputs. In fact, they possess gain spectra that are intrinsically broad, and beyond that can even be tailored to cover an octave-spanning range.

This thesis explores the development of broadband sources of terahertz radiation based on quantum cascade lasers (QCLs). The chief way this is done is through the development of compact frequency combs based on THz QCLs, which are able to continuously generate milliwatt levels of terahertz power covering a fractional bandwidth of 14% of their center frequency. These devices operate on principles similar to microresonator-based frequency combs, and make use of the quantum cascade laser's fundamentally large nonlinearity to phase-lock the cavity modes. These devices will enable the development of ultra-compact dual comb spectrometers based on QCLs, and will potentially even act as complete terahertz spectrometers on a chip.

This thesis also uses broadband terahertz time-domain spectroscopy to analyze the behavior of THz QCLs. By using QCLs as photoconductive switches, the usual limitations imposed by optical coupling are circumvented, and properties of the laser previously inaccessible can be directly observed. These properties include the gain and absorption of the laser gain medium, the populations of the laser's subbands, and properties of the waveguide like its loss and dispersion. Knowledge of these properties were used to guide frequency comb design, and were also used to inform simulations for designing better lasers.

Thesis Supervisor: Qing Hu
Title: Professor

Acknowledgments

First of all, I would like to thank my thesis advisor, Professor Qing Hu, for providing me with the opportunity to work on this project. His enthusiasm for his research is obvious to anyone, and the fact that his group has a core technology like terahertz quantum cascade lasers is extremely helpful for students who want to do something impossible in any other setting. I am especially grateful for his patience and willingness to allow me to go out on a limb even when the outcome was far less than certain.

I would also like to thank Professors Erich Ippen and Keith Nelson for taking the time to serve on my thesis committee, and also Franz Kaertner and Noah Chang for having valuable discussions with me that got me interested in frequency combs. I also owe much to Professor Dayan Ban of Waterloo for initiating my terahertz-time domain project and giving me a great starting point for my terahertz research. Special thanks also goes to Dr. John Reno at Sandia for providing excellent MBE growth and to Drs. Jian-Rong Gao and Darren Hayton of SRON for lending me one of their best HEB mixers and telling me how to use it.

It almost goes without saying that I have been helped by a huge number of people in the lab during my time here. In particular, Wilt Kao was pretty much everything I could ask for in a peer, always willing to converse about research and life with me and pushing me to be my best. He also fabricated nearly all of my devices, so I owe a great deal of this thesis to him. Alan Lee is probably the most well-rounded person I met at MIT and is pretty much great at everything. He was an excellent role model who is always willing to do whatever it takes for his research and doesn't shy away from dirty work. Ningren Han impressed me with his resourcefulness and has continued to be a great friend since leaving the lab, offering valuable advice on many subjects. It was his intervention that netted me the journal cover I so enjoy.

My interactions with others in the lab have been similarly rewarding. Ivan Chan is one of the most helpful people I know and was always willing to hear me out on my various and sundry ideas about physics and about QCL design. Qi Qin showed

me how to work efficiently towards a goal and always had an interesting perspective on many topics. Shengxi Huang is one of the most persistent people I've ever met and seemed to know everything about everything and everyone. Allen Hsu and Sushil Kumar were extremely helpful for teaching me lab technique early on in my graduate career. Xiaowei Cai and Yang Yang were excellent office-mates during my last couple years and even helped me with experiments despite their junior status. Last but not least, I'd like to thank my other labmates Amir Tavallaee, Xinpeng Huang, Indrasen Bhattacharya, Asaf Albo, Ali Khalatpour, and David Levonian for providing intellectually-stimulating discussion.

Finally, I'd like to thank my parents for raising me in a household that emphasized the importance of education. I'd also like to thank my brother, Dan, for being a terrific lifelong friend, and my sisters, Laura, Emily, and Jenny, for always looking up to me and giving me support.

Contents

List of Figures

Chapter 1

Introduction

In recent years, terahertz quantum cascade lasers (THz QCLs) have proven to be an effective compact source of continuous-wave terahertz radiation, defined here as the frequency range spanning 1 to 10 THz [1]. In a sense, terahertz light represents the "final frontier" of non-ionizing radiation, given that its generation has traditionally been inaccessible to conventional sources. Fast electronic sources such as high electron mobility transistors [2], heterojunction bipolar transistors [3], and Gunn diodes [4] can be used to generate radiation from DC to microwaves, but parasitic roll-offs limit their frequency response to a few hundred gigahertz. Traditional interband semiconductor diode lasers can operate in wavelengths ranging from as short as the UV to as long as the mid-infrared [5], but are limited to quasi-visible wavelengths by the bandgap of the material. Figure 1-1 shows this phenomenon, also known as the "terahertz gap" [6].

As a result of this gap, the applications for this field are not entirely well-known and it remains severely underdeveloped, at least in comparison with other frequency ranges. Though many applications have been proposed in the literature, to some extent terahertz has been oversold by researchers in the field, and there isn't yet a "killer app" for terahertz technology. For example, homeland security is frequently mentioned as an application as a result of terahertz's ability to peer through clothing while being non-ionizing [9], but millimeter-wave scanners are already in place in most airports and are less susceptible to effects like atmospheric absorption. Like-

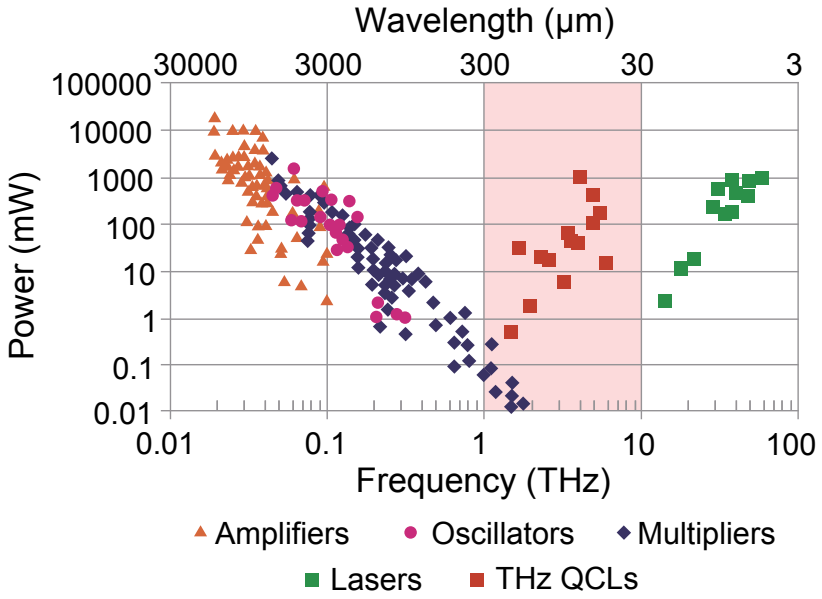


Figure 1-1: Frequency and power of various terahertz sources. From Refs. [7] and [8].

wise, non-invasive detection of explosives through clothing has to overcome many challenges in order to be viable [10]. A better application is the use of terahertz to non-destructively monitor industrial processes involving materials that are optically opaque but terahertz transparent, such as pharmaceutical coatings [11] and space shuttle foam [12]. Though these applications are certainly valuable, they are limited in scope and fairly niche. Researchers also tout terahertz for its theoretical ability to form high-bandwidth line-of-sight communications links [13, 14, 15], but it is hard to see how terahertz might out-compete telecommunications wavelengths in this regard, especially when the relative difference in technological development is considered.

1.1 Terahertz spectroscopy and frequency combs

Out of all the potential applications, spectroscopy seems to be the most promising, since it is the only application that can make use of the properties unique to terahertz. More specifically, since many molecules have strong rotational and vibrational resonances in the terahertz regime [16, 17], spectroscopy at these wavelengths can

elucidate structural changes. For example, Figure 1-2 shows the structure of various explosives, in addition to their fingerprints in the terahertz. The most common

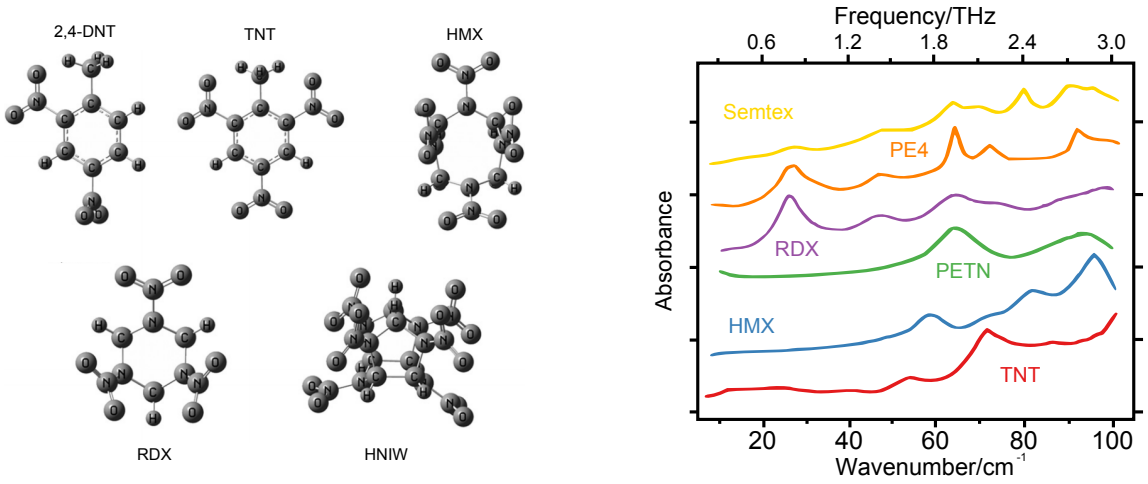


Figure 1-2: Chemical structure of various explosives (from Ref. [18]), as well as the corresponding terahertz absorption spectra (from Teraview).

way of performing spectroscopy at these wavelengths is a technique known as terahertz time-domain spectroscopy (THz TDS). Essentially, the output of a near-infrared mode-locked laser is downconverted to terahertz, and the resulting field is sampled by the same mode-locked laser. Although many people have been able to achieve excellent results within laboratory environments using this technique, it has not been able to achieve wide commercial viability since it requires a mode-locked laser and an optical setup.

Another way of performing terahertz spectroscopy is in the frequency domain, using laser spectroscopy. A narrow-linewidth laser (typically a distributed-feedback THz QCL) is tuned across the absorption feature of interest, and the power of the laser is measured. An example spectrum is shown in Figure 1-3. This technique is particularly well-suited for high-resolution spectroscopy, but is limited by the tuning range of the laser. In addition, frequency calibration is extremely challenging: without a frequency reference, one must rely on spectral databases like HITRAN to figure out exactly what frequency is being measured.

If compact terahertz spectrometers could be made, that would significantly im-

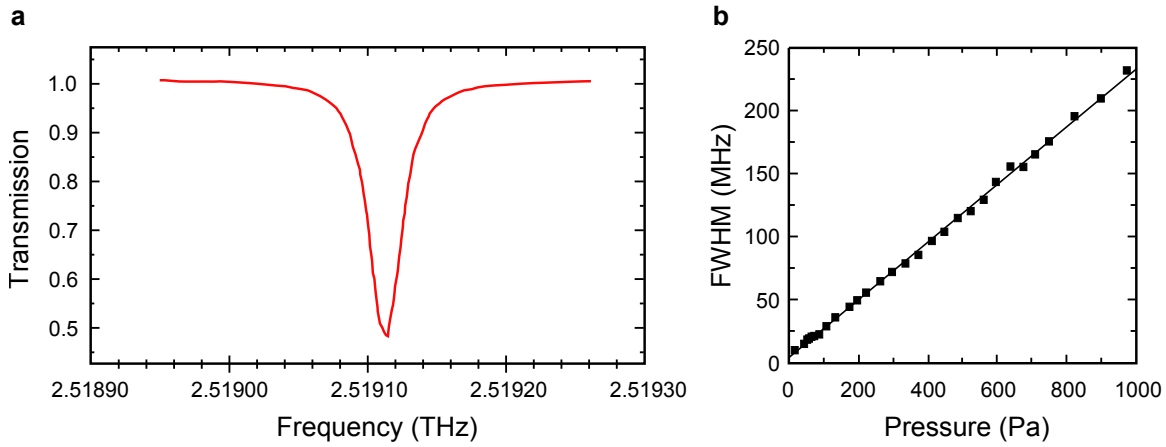


Figure 1-3: Absorption of methanol at 100 Pa, measured using a THz QCL. From Ref. [19].

prove terahertz's ability to move out of the lab. Any spectroscopic system requires three basic elements: a source, a detector, and a way of performing frequency discrimination. Although people frequently place most of their attention on the sources and detectors, it is often the case that frequency discrimination is what limits the compactness of a spectrometer. This is due to the fact that the best ways of performing frequency discrimination—grating spectrometers and Fourier Transform Spectrometers—typically require a physical delay element to achieve optimal levels of performance. Even at the extremely technologically well-developed visible wavelengths, there is no such thing as a spectrometer smaller than a few centimeters that can achieve performance equivalent to an optical spectrum analyzer or an FTIR[1] in terms of resolution and throughput.

Optical frequency combs, on the other hand, offer a way to make compact spectrometers. A frequency comb—shown in Fig. 1-4—is a light source whose lines are evenly spaced and well-defined. The lines of a frequency comb can be characterized by two parameters: an offset ω_0 and a repetition rate $\Delta\omega$. As such, they effectively act as a ruler for optical frequencies, and as a result have revolutionized the fields of precision metrology and spectroscopy. At visible and near-infrared wavelengths,

[1]Fourier Transform Infrared Spectrometer

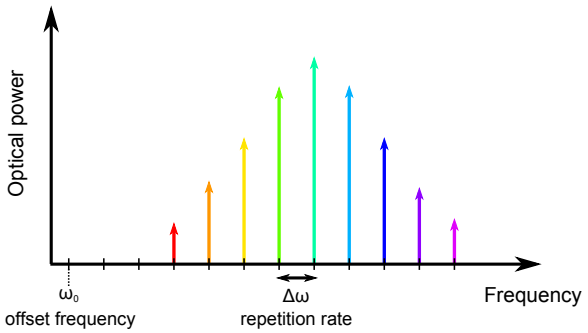


Figure 1-4: Basic frequency comb.

the frequency comb can be used to build ultra-precise clocks, to probe the fundamental physics of atomic transitions, or even to infer the presence of extrasolar planets. The reason frequency combs are particularly tantalizing for compact spectroscopy is that they can be used to replace the mechanical delay element typically necessary for a high-resolution spectrometer. Moreover, through a technique known as dual-comb spectroscopy, they can actually perform *better* than delay elements, easily providing resolutions equivalent to delays of kilometers or better. Also known as multi-heterodyne spectroscopy, the principle is simple and is shown in Figure 1-5. Two combs with slightly different repetition rates are shined onto a single fast

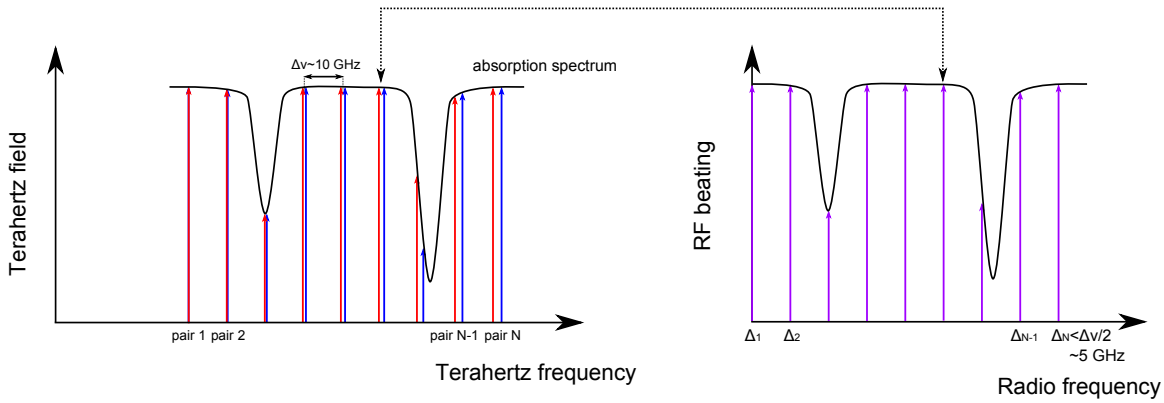


Figure 1-5: Dual comb spectroscopy. Two combs with different repetition rates generate RF beatnotes that encode the optical spectrum.

detector. Because the beating between adjacent lines is different for each pair, the resulting beatnotes are unique, with their magnitudes encoding the geometric mean

of the intensity of the two lines constituting the beatnote. If one or the other is shined through a sample, these beatnotes uniquely encode the absorption of that spectrum. Of course, the measured spectrum is discontinuous, but if the offset frequency of one comb or the other can be swept then a full gapless spectrum can be measured.

It is worth mentioning that pulsed THz sources are, in fact, frequency combs. They are pulse trains in the time-domain with well-defined phases, and are therefore combs in the frequency domain. In fact, asynchronous optical sampling—a technique in which two mode-locked lasers have their repetition rates detuned from each other to perform THz TDS—is very similar to dual-comb spectroscopy [20]. However, pulsed sources have difficulty achieving powers of more than a few microwatts and are typically detected coherently. Moreover, they are bulky and involve expensive mode-locked lasers. THz QCLs, on the other hand, are powerful compact sources of such radiation capable of generating Watt-level power output. If they could be made to generate the uniformly-spaced lines of a frequency comb, they would be ideal candidates for making compact spectrometers. Until this work, no one had been able to generate THz QCL combs with more than a few lines, so one of the major goals of this thesis is to demonstrate how this can be done.

1.2 Quantum cascade laser basics

QCLs are created by periodically growing alternating layers of different materials on a semiconductor substrate, such as GaAs and $Al_{0.15}Ga_{0.85}As$. The layers are grown with molecular beam epitaxy (MBE), and because each layer is only a few monolayers thick, quantum-size effects dominate and create an artificial band structure. Thus, the designer can tailor the wavelength of the structure to a wide range simply by changing the growth thicknesses. As a result, QCLs have no theoretical upper bound on the wavelengths they can generate, giving them tremendous versatility for implementing long-wavelength oscillators. Because each QCL period (known as a module) can be repeated hundreds of times, each electron injected into the system can emit hundreds of photons. Moreover, since the confinement of each electron

is one-dimensional, the conduction band splits up into a series of two-dimensional subbands, each corresponding to an artificially-designed state. Since the dispersion relation of each subband is approximately equal, all momentum-conserving transitions will occur at the same energy. In other words, the so-called joint density of states is a delta function, and every electron in a subband can contribute to lasing. Again, this is in contrast to bulk semiconductor lasers, in which only a small fraction of the electrons in the system are at the same energy and can contribute to gain [21].

Historically, the first QCL was demonstrated in 1994 in the mid-infrared at Bell Labs [22], while the first terahertz QCL was demonstrated in 2002 at the Scuola Normale Superiore [1]. The reason for this long delay is that there were several challenges unique to terahertz, requiring that different design methodologies be used. First, though mid-infrared QCLs had been developed using conventional dielectric waveguiding to confine the optical mode, this technique cannot be used in terahertz QCLs due to the impracticality of growing a wafer whose gain medium thickness (10 µm) is on the order of a wavelength (100 µm). Therefore, sub-wavelength plasmonic techniques are required to effectively confine the mode. This problem was essentially solved by development of semi-insulating surface plasmon waveguides [1], which confine the mode between a layer of metal and a highly-doped plasmon layer, and also by the development of metal-metal waveguides [23], which confine the mode between two metal layers in a method similar to microstrip lines. These waveguides are shown in Fig. 1-6.

The second major challenge associated with terahertz QCLs is the design of the gain medium itself. As with practically any laser, the primary goal when designing QCL gain media is to achieve population inversion, through the selective population of the upper laser state and the selective depopulation of the lower laser state. Selective population is achieved in most mid-infrared and terahertz QCLs with the use of a set of quantum wells known as an injector, which is separated from the upper laser level through a barrier called the injector barrier. The injector barrier is thick enough that the only way electrons can get into the next module of the system is by resonantly tunneling through the barrier into the upper state, preventing other states

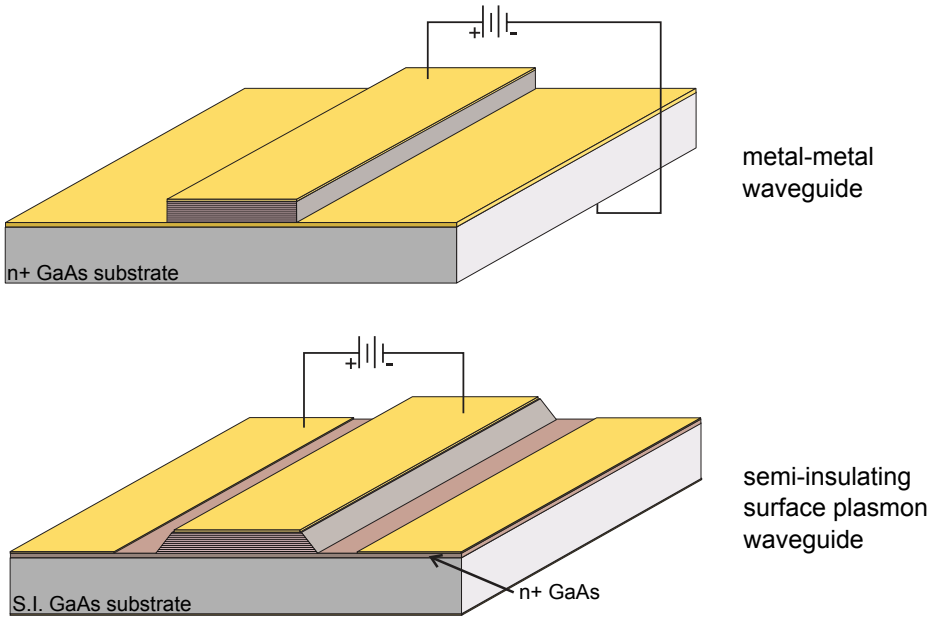


Figure 1-6: Terahertz waveguides used for THz QCLs. From Ref. [24].

from accidentally being populated. While this works to provide selective upper state population for both mid-infrared and terahertz designs, depopulation of the lower laser state is unfortunately not so simple. In mid-infrared QCLs, a convenient way to depopulate the lower state is to separate it from another state by the energy of a longitudinal-optical (LO) phonon, about 36 meV in GaAs. [25] Because all LO phonons have approximately the same energy in bulk material, LO phonon scattering can be extremely fast (sub-ps), and the lower state remains depopulated. However, for terahertz QCLs, this design is problematic. In order for the gain of the system to be large, the spatial overlap between the upper and lower laser states should be substantial, which means that the spatial overlap between the upper laser level and the LO phonon-separated level will also be important. As a result, not only will electrons be able to efficiently scatter from the lower laser level into the LO-phonon level, but they will also be able to scatter nonradiatively from the upper laser level into the LO-phonon level. This parasitic scattering channel reduces population inversion.

In order to get around this issue, several types of depopulation schemes have been created, among them the chirped-superlattice [1] and bound-to-continuum designs

[26], which utilize a miniband of tightly-coupled states coupled to the lower laser level. The one pioneered by the MIT group is known as a resonant-phonon design [27], and operates by resonantly coupling a state known as the collector to the lower laser level, which is in turn separated from another state by an LO-phonon energy. Figure 1-7 shows a prototypical resonant-phonon design. State 1' is the injector, state 4 is the upper laser level, state 3 is the lower laser level, state 2 is the collector, and state 1 is another injector state. In this design, an electron in the injector first tunnels resonantly into the upper laser state, where it emits a photon into the lower laser state. It resonantly tunnels again into the collector state, and finally emits an LO phonon into another injector state.

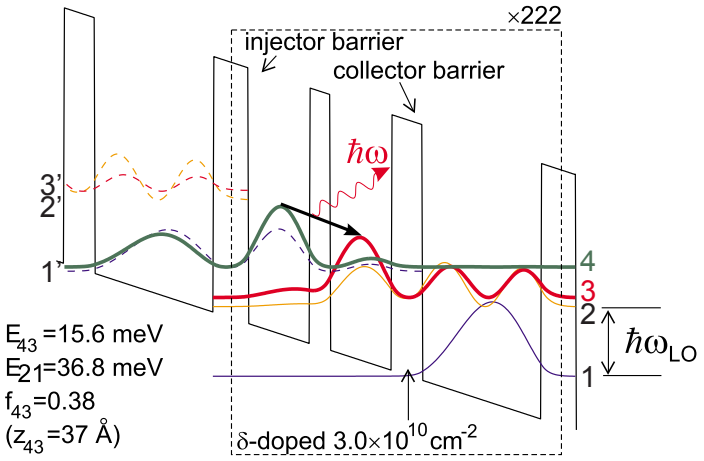


Figure 1-7: Resonant phonon QCL that operates at 186 K. From Ref [28].

1.3 Gain medium spectroscopy

However, even though terahertz QCLs were demonstrated long ago, they are still in many ways deficient compared to their mid-infrared counterparts, primarily in the realm of temperature performance. Mid-infrared QCLs have long been able to operate at temperatures exceeding 425 K [29], but terahertz QCLs are currently limited to a maximum temperature of 200 K. [30] Phenomenologically speaking, one explanation for this limitation is that when the photon energy $\hbar\omega$ is approximately equal to kT,

maintaining inversion will be difficult due to thermal backfilling and other effects. Since no GaAs QCL can operate at frequencies far above 5 THz due to the strong absorption of the LO-phonon line, this rubric would limit THz QCL operation to a maximum temperature of about 240 K. Unfortunately, this temperature necessitates the use of cryogenics, severely limiting potential applications and increasing costs. As a result, one of the major goals in the field right now is to achieve room temperature performance. Figure 1-8 shows a summary of the maximum operating temperature of different published designs, along with a line indicating $\hbar\omega/kT = 1$.

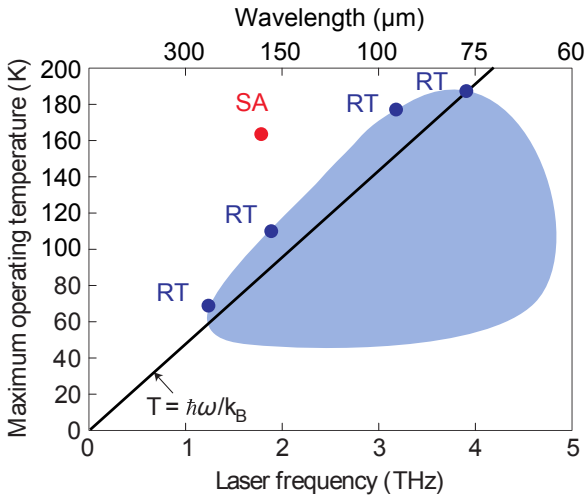


Figure 1-8: Summary of the published THz QCL designs' temperature performance. The shaded region shows designs that have were published as of January 2011, while RT refers specifically to devices which use a resonant-tunneling injection mechanism. From Ref. [31].

Note that while the $\hbar\omega \sim T_{max}$ rule of thumb generally holds true, it is not absolute, and the device in Ref. [31] managed to significantly exceed this limit by utilizing a nonstandard injection mechanism. There are two major arguments for why injection-related effects are believed to play an important role in temperature degradation of lower-frequency QCLs. First, because low frequencies are closely-spaced in energy, design requires that the injector barrier be made relatively thick in order to reduce parasitic coupling of the injector to the lower state. Making the injector barrier thicker also makes the tunneling process more incoherent, thereby reducing the maximum current that can be pushed through the structure and reducing

the available gain [32]. Secondly, a reduction in frequency results in a reduction of the dynamic range of the structure, due to the formation of negative differential resistance (NDR). NDR is a well-known phenomenon that occurs in many systems and results from the misalignment of two subbands that are resonantly-coupled. Once two subbands are maximally-aligned, increasing the voltage bias will cause them to misalign, decreasing the current flow and causing NDR. In THz QCLs, this is a problem because the NDR creates an instability in the voltage bias across the structure which causes the bias field to break up into multiple domains, making the device behave somewhat unpredictably. Stable device operation can therefore only happen in the regions of positive resistance, that is, at biases where the injector state is between the upper and lower laser levels. Since this range is proportional to the photon energy, the dynamic range and therefore maximum temperature of operation will be a linear function of frequency [31]. Another major factor believed to contribute to the temperature degradation of THz QCLs is the effect of thermally-activated LO phonon scattering. Even though the energy difference between subbands is below E_{LO}=36 meV for THz QCLs, electrons within the subband are thermalized at a temperature significantly higher than the lattice temperature [33]. Sufficiently hot electrons are separated from states of the lower subband by E_{LO} and can therefore emit a phonon, thereby reducing the available population inversion [6]. Note that this process actually favors lower-frequency designs, because more thermal energy is necessary to "activate" the LO phonon emission.

Unfortunately, despite the effort of many theorists and experimentalists, temperature performance of THz QCLs seems to have stalled, and no improvement has been made in over two years. One explanation for this is that the large number of processes involved in QCL electron transport makes the system somewhat difficult to model. A full simulation needs to take into account coherent transport (i.e., tunneling), pure dephasing, LO phonon scattering, impurity scattering, electron-electron scattering, interface roughness, and more [6]. To make matters worse, some of the properties are MBE-growth dependent, chiefly the interface roughness. The simplest approach is to use rate equations, calculating scattering rates from Fermi's Golden Rule, but this

cannot take into account the effect of coherence. For example, this approach would predict that tunneling current through a barrier should effectively be instantaneous and not dependent on barrier thickness, even though this is not the the case. The next level of sophistication is to include the effects of coherence by keeping track of density matrices, whose diagonal elements represent the populations and whose off-diagonal elements represent coherences. Once again, the lifetimes are calculated numerically, but here a phenomenological dephasing term is added to the time-evolution of the density matrix that effectively "localizes" the wavefunctions. Gain can then be calculated numerically, using a Monte Carlo approach [32], or semi-analytically, which is helpful for gaining an intuitive understanding of QCL behavior [34]. The most complete approach is to use one based on non-equilibrium Green's functions, which calculates electron transport essentially from first principles [35, 36, 37]. Though this approach has become more practical in recent years with advancing computing power, it is very complex and is usually performed by dedicated theorists. Ultimately, the most successful approach to QCL design has been a semi-quantitative approach, combined with a large number of design iterations.

Experimentally speaking, QCL design has been hindered by the inability to measure many of the degrees of freedom involved in electron transport. The most straightforward measurement—the light-voltage-current relationship—essentially tells how well a laser performs, but not very much about the physics. Conductivity measurements provide a little more information about which states are aligned, but not much else. The "holy grail" of QCL analysis would be able to measure electron populations in a way that allows all of the states of the system to be probed. Gain medium spectroscopy is one way to do this: by measuring the absorption and gain of a device under bias, one can in principle extract the populations and linewidths of every populated transition. In mid-infrared QCLs, this has been done by simply coupling broadband radiation from an incoherent blackbody source into the facet of a QCL, collecting the radiation emitted from the other side, and comparing the intensity of the emitted light in each polarization [38]. Since the intersubband transitions of a QCL only affect light polarized in the MBE growth direction, this means that an ab-

sorption profile can be deduced. However, this is problematic in the terahertz: since the emissivity of a blackbody scales with frequency squared at low frequencies, the terahertz emission is weak, even at high temperatures.

An alternative technique available in the terahertz range is the use of time-domain spectroscopy to probe THz QCLs, first demonstrated in 2007 [39]. The concept is similar to the mid-infrared transmission experiment—sending broadband radiation through a laser ridge in order to measure its absorption and gain—but the source is very different. A broadband terahertz pulse is instead generated using an optical pulse from a mode-locked laser, the resulting pulse is propagated through a QCL, and the final pulse is sampled by a delayed version of the original optical pulse. By scanning the delay, a temporal profile of the propagated terahertz pulse is generated, giving access to both its amplitude and phase. When the pulse is compared to a reference pulse obtained with the QCL off, the gain/absorption induced by bias can be found.

This method can be used to do many types of classical laser experiments, including showing the presence of gain clamping [39] and spatial hole burning [40]. However, there has been to date very little quantitative information gleaned from these types of measurements about the gain medium itself that couldn't have been obtained from other types of measurements. For example, the linewidth of the lasing transition can be found by measuring electroluminescence from a device biased below threshold, while the cavity losses can be measured by simply comparing devices of different lengths. The core reason for this is that in all TDS experiments performed until recently, the semi-insulating surface plasmon (SISP) waveguide has been tested rather than the metal-metal waveguide, due to its superior in- and out-coupling efficiency. The metal-metal waveguide tightly confines its mode to the 10 µm active region, while the SISP waveguide allows its mode to extend significantly into the substrate. As a result, the impedance mismatch of metal-metal waveguide modes and free space is quite high, while the mismatch of an SISP waveguide is similar to that of bulk GaAs. Usually, this property of metal-metal waveguides is considered advantageous, since the high mismatch creates a high cavity mirror reflectivity ($R \sim 0.85$), thereby

reducing the mirror losses and improving the temperature performance [41]. However, this poses a problem when the device is probed with a terahertz pulse, since a very small fraction of the pulse will couple in and out of the waveguide. To make matters worse, the signal will generally be swamped in stray light, light that never actually entered the cavity but was collected because it was in proximity.

Though reasonable results can be obtained with transmission-mode SISP devices, they are suboptimal for analyzing the physics of a gain medium. As SISP devices have mode profiles with a relatively small overlap with their gain medium, this reduces the gain medium's interaction with terahertz pulses. Stray light is still a problem with these waveguides, but its effect can be reduced by attaching a metallic aperture to the laser ridge. Unfortunately, no aperture is perfect, and the stray light collected reduces the magnitude of any features on the gain spectrum. Most problematic is the fact that all of the QCLs with the best temperature performance have metal-metal waveguides, a result of their reduced mirror losses and better mode confinement. Practically speaking, there is very little reason to use valuable MBE growth resources to make SISP wafers, especially when their only purpose is to be probed.

Until this work, no one had been able to perform spectroscopy on QCLs with metal-metal waveguides. By using QCLs as photoconductive switches, the usual limitations imposed by optical coupling were circumvented, and the highest dynamic-range spectroscopic measurements of QCLs were performed. These techniques enabled direct observation of properties such as the gain and absorption of the laser gain medium versus temperature, the populations of the laser's subbands, and properties of the waveguide like its loss and dispersion. It was also used to study the coherence of electron transport and the lineshapes of various designs. To our knowledge, this has never been directly observed in a structure based on quantum wells alone. Knowledge of these properties were used to guide frequency comb design, and were also used to inform simulations used to design better lasers. These simulations were used in genetic algorithms to design lasers with the goal of increasing temperature performance.

1.4 Thesis goals and organization

This thesis is organized in such a way that each chapter represents a major experimental goal.

- Chapter 2 establishes how integrated emitters were used to probe THz QCLs using terahertz time-domain spectroscopy.
- Chapter 3 contains the majority of the gain medium characterization that was performed with THz-TDS.
- Chapter 4 discusses the key issues involved in designing THz QCL based frequency combs, as well as the early results that suggested comb formation.
- Chapter 5 contains the coherence measurements that were used to definitively establish that the lasers designed to be combs had the coherence properties of combs.

There are also several appendices that discuss topics which can be decoupled from the main text and serve as useful references.

- Appendix A summarizes the periodic density matrix formalism that was developed to model the bias- and frequency-dependence of QCL gain.
- Appendix B lists the electrical modulation schemes that were used for time-domain measurements.
- Appendix C derives in more detail the underpinnings of SWIFTS, the key technique used to characterize comb coherence.

Chapter 2

Terahertz time-domain spectroscopy of quantum cascade lasers

Terahertz time-domain spectroscopy (THz-TDS) is the standard technique for generating and detecting broadband terahertz radiation. Although its resolution is limited compared with laser-based spectrometers, its high signal-to-noise ratio and broadband coverage made it superior to earlier techniques like Fourier transform spectroscopy. Developments in the field include the creation of photoconductive switches in the 1980s [42], the development of $\chi^{(2)}$-based techniques like optical rectification and electro-optic sampling [43] in the 1990s, and the advancements made in air-biased coherent detection in the 2000s [44]. Because THz-TDS is so useful for probing samples, it is also extremely useful for characterizing THz QCLs, both the gain media and the waveguides. This chapter reviews the basic principles of THz-TDS and discusses how it was used to characterize THz QCLs.

2.1 General principles of THz-TDS

Figure 2-1 shows a block diagram of a typical THz-TDS system, as well as a schematic of how the data is recorded. THz-TDS is ultimately a pump-probe technique, and is based on near-infrared optical pulses from a pulsed laser, typically a mode-locked Ti:Sapphire or a fiber laser. The pump pulse is shined onto a source and generates

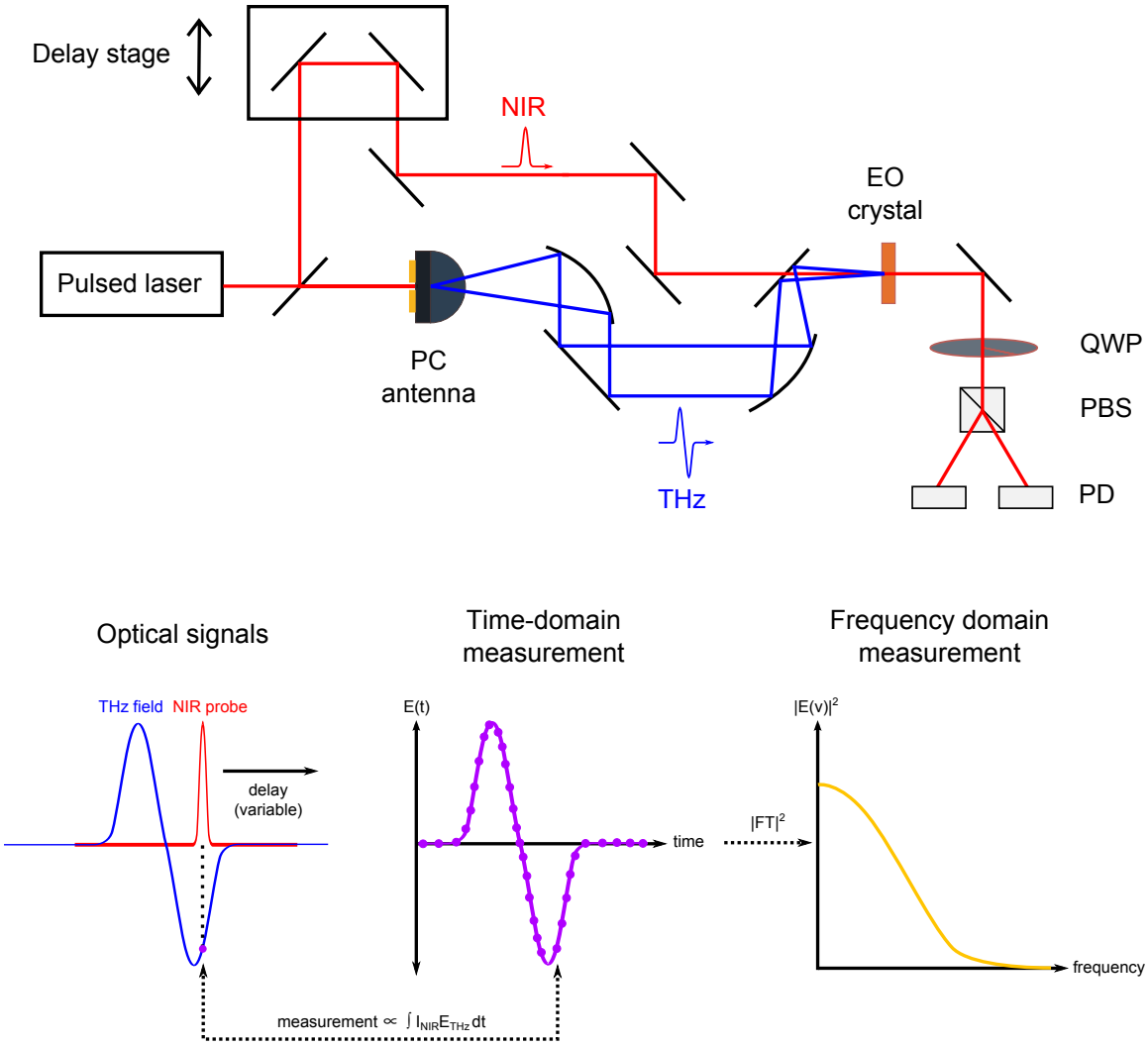


Figure 2-1: Typical TDS system. A sample can be placed in the THz beam path to measure its transmission.

broadband terahertz pulses. (The source shown here is a photoconductive antenna, discussed further in section 2.1.1.) The terahertz light that is produced and is re-combined with a delayed version of the probe pulse, and the two are mixed in the detection element. (The detector shown here is electro-optic, discussed in section 2.1.2.) If the resulting signal is proportional to the product of the THz field and the near-IR intensity, by scanning the delay stage one can measure the terahertz field on sub-picosecond timescales, limited only by the width of the probe pulse. Since the entire electric field waveform is obtained, Fourier transforming the result reveals both

the amplitude and the phase of the terahertz pulse as a function of frequency.

Of course, any real system will have many important practical factors which enter into its design. Some of these considerations are discussed in the following subsections, but they are by no means comprehensive.

2.1.1 Sources

An interesting fact about ultrafast science is that many materials will respond to ultrashort optical pulses by producing terahertz waves. For example, if an undoped semiconductor is illuminated with an ultrashort pulse, its electrons will absorb the radiation and be promoted from the valence band to the conduction band. This temporarily increases the conductivity of the sample, and if it is biased a transient current will be generated. This current radiates in the terahertz, and so the resulting device is called a *photoconductive switch.* Similarly, for crystals that possess $\chi^{(2)}$ nonlinearity, because ultrashort pulses possess a broad banndwidth, difference frequency generation can occur that generates broadband radiation through a process colloquially referred to as *optical rectification* [45]. Even unbiased semiconductor surfaces can generate terahertz through a process known as the photo-Dember effect [46, 47], provided that the diffusion constant of electron and holes is sufficiently different to allow for a local space charge to form upon optical excitation (as they are in light-electron materials like InAs).

The sources used in this thesis for THz-TDS are all photoconductive in nature, so more discussion of them is in order. A basic photoconductive antenna is shown in Figure 2-2. Gold contacts are patterned onto an epitaxial layer of GaAs grown at 200°C (known as low-temperature or LT-GaAs), which is chosen for its extremely short carrier lifetime of ~0.4 ps that leads to a broad terahertz response [48]. The contacts also function as an antenna for the radiation that is generated, with different shapes providing different frequency responses. For example, the bowtie shape shown here will be fairly broadband owing to its self-similarity, whereas a dipole-type antenna will be enhanced at its resonant frequency.

To use the antenna, light from a Ti:Sapphire near 800 nm is shined between the

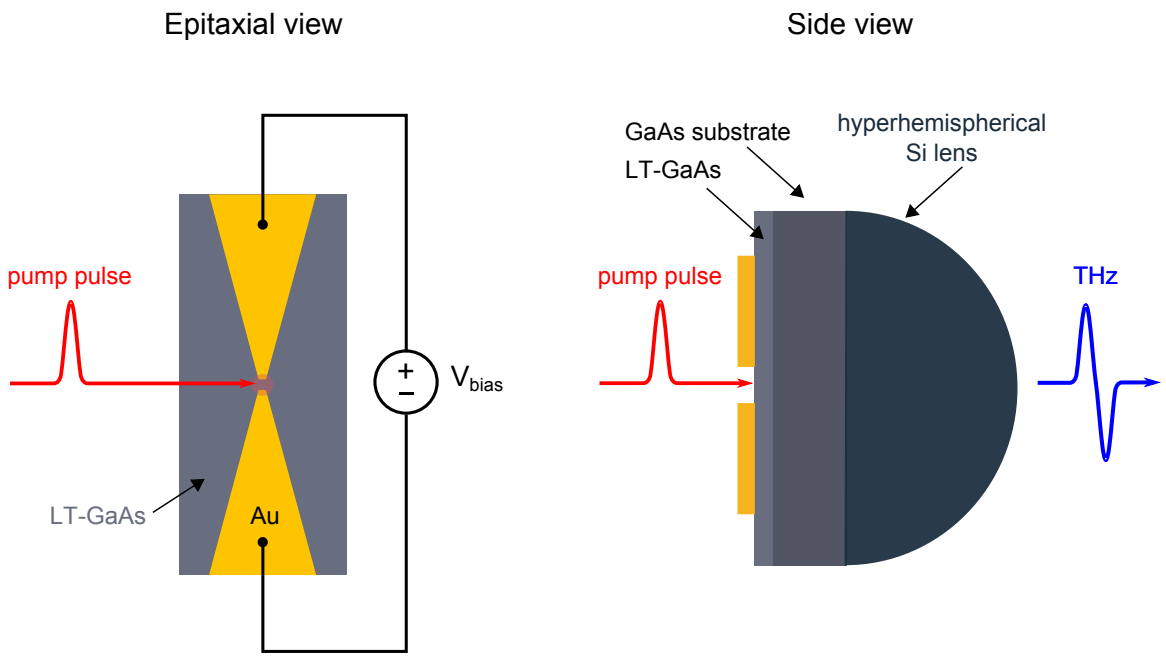


Figure 2-2: Schematic of a basic photoconductive antenna based on LT-GaAs.

contacts, and the current transient generated by the temporarily excited free carriers radiates in the terahertz. If the excitation is smaller than the terahertz wavelength (which is typically the case as a diffraction-limited spot size on the order of 10 microns will be much smaller than the wavelengths accessible to TDS), it will radiate in all directions. However, because the substrate has a higher refractive index than air, most of the power of the antenna radiates preferentially into the substrate, with a ratio of $1 : n^3 = 46$ for GaAs [49, 50, p. 9]. Therefore, terahertz light is collected from the substrate side, usually with the aid of a high-resistivity silicon lens. This lens has two effects. First, it changes the geometry of the system to allow most of the rays emanating from the antenna to hit the surface of the lens at a shallower angle, thereby preventing the total internal reflection that would otherwise trap the light in the substrate. Secondly, if the thickness of the substrate plus lens system is chosen to be $R + R/n$ (i.e., the setback from the center of the sphere is R/n), then the system will be aplanatic and will converge the rays coming from the center of the antenna without spherical aberration [51]. This reduces the numerical aperture that is required of any collection mirrors and greatly increases the amount of light that

can be out-coupled from the antenna.

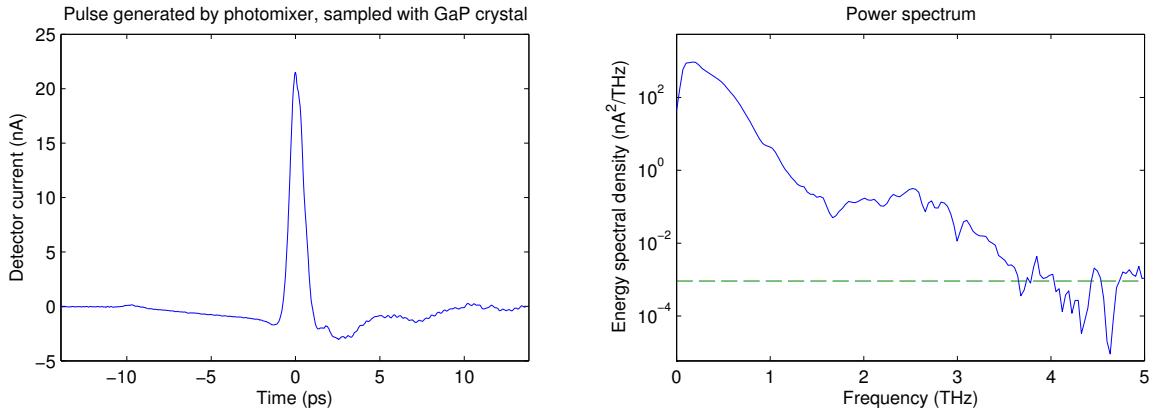


Figure 2-3: Pulse generated by an LT-GaAs photoconductive antenna using a bowtie geometry, along with its power spectrum.

Figure 2-3 shows an example of a terahertz pulse that can be generated by a bowtie-type PC antenna like the one in Figure 2-2, along with its power spectrum (as measured by electro-optic detection, discussed in section 2.1.2). The dotted line represents the noise floor of the measurement. Notice first that in the time domain, the electric field waveform resembles a delta function which has been band-pass filtered. This is because the radiated electric field of the antenna will be related to the temporal derivative of the antenna current and therefore the derivative of the instantaneous carrier density, $n(t)$:

$$E(t) \propto \frac{\partial J}{\partial t} \propto \frac{\partial n}{\partial t}. \tag{2.1}$$

If the carrier density is an ideal step-function (turning on as soon as the pulse arrives and decaying slowly thereafter), then the electric field waveform will be a delta function. Of course, practical limitations limit the bandwidth that can be collected. The low-frequency ($<$100 GHz) components have been removed because the bowtie has a finite size and by definition cannot radiate into the far field. The high-frequency limitations of the measurement typically limited by the width of the pump pulse, the transit time associated with traveling the gold contacts (leading to RC roll-off), and the absorption of the GaAs substrate and the hyperhemispherical lens. It is possible to generate bandwidths as high as 30 THz with PC antennas (as demonstrated by

Shen et al. and shown in Figure 2-4), but this requires that the terahertz be collected in a reflection-type geometry. This result also necessitates the use of short optical pulse (approximately 15 fs).

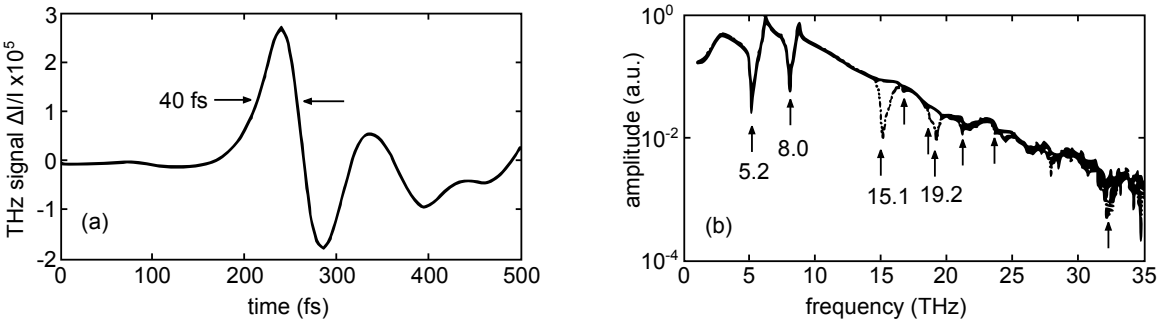


Figure 2-4: Time- and frequency-domain waveforms obtained from a photoconductive antenna whose backwards-generated radiation was collected instead of its forward-propagating radiation. From Ref. [52].

For typical commercial Ti:Sapphire systems (such as the Spectra-Physics Tsunami used for most of this work, which reliably produces 70 fs pulses at a repetition rate of 80 MHz at a wavelength of 780 nm), photoconductive antennas usually produce higher output power than rectification-based sources, at the expense of terahertz bandwidth. They are also fairly simple to operate, as they work at room temperature, have no phase-matching requirements, and only require that the pump pulse be shined on the active element. However, unlike rectification-based sources they cannot be scaled with the intensity of the pump pulse, saturating once the optical power is high enough to fully short the contacts.

2.1.2 Detectors

Though this thesis focuses on terahertz sources, terahertz detection is still a whole area of study in and of itself. Terahertz time-domain pulses can naturally be directly detected by electronic means such as Schottky diodes [53], by direct detectors such as gallium-doped germanium detectors [54], or even by bolometric action as in a hot-electron bolometer [55]. Having said that, such methods of detection are all poor ways of detecting time-domain pulses, since the typical time-domain pulse has an

extremely short duty cycle. For example, a pulse that is 1 ps long and is generated by a laser with an 80 MHz repetition rate will have a duty cycle of less than 0.01%. As direct detection integrates noise from all times, the signal-to-noise ratio (SNR) of such a measurement will be unnecessarily reduced by a factor of $1/\sqrt{\mathrm{DC}}$, which is more than 100 in the case of an 80 MHz Ti:Sapphire.

It is for this reason that practically all THz time-domain systems utilize a technique known as coherent detection. The key idea is to sample the THz electric field by the delayed probe pulse, thereby ensuring that the system is only sensitive to noise that occurs at the same time as the probe pulse. Coherent detection massively increases the SNR of TDS systems, and is what makes them truly practical. Two of the most common ways of performing coherent detection—photoconductive sampling and electro-optic detection—are essentially the inverse processes associated with the previously-discussed generation processes. In photoconductive sampling, the probe pulse and terahertz radiation are both shined onto an unbiased element similar to a photoconductive antenna. The probe pulse temporarily increases the conductivity of the semiconductor, and the terahertz field sweeps carriers to one of the contacts. This induces a DC current that can be detected with a lock-in amplifier. Note that the current detected is proportional to the field "seen" by the probe pulse, that is the instantaneous terahertz field.

Electro-optic (EO) detection, the method primarily used for coherent detection in this thesis, is the inverse process of optical rectification [56, 57, 58, 59]. Essentially, it works by detecting the polarization rotation induced in the probe pulse by the terahertz field in an electro-optic crystal like ZnTe, GaP, or GaAs. What these materials have in common is that the application of an electric field $\mathbf{E}$ induces an anisotropy in their permittivity tensor, which due to their zincblende structure takes the form

$$\epsilon_{\mathbf{r}}{}^{-1}(\mathbf{E}) = \begin{pmatrix} \frac{1}{n^2} & 0 & 0 \\ 0 & \frac{1}{n^2} & 0 \\ 0 & 0 & \frac{1}{n^2} \end{pmatrix} + r_{41} \begin{pmatrix} 0 & E_z & E_y \\ E_z & 0 & E_x \\ E_y & E_x & 0 \end{pmatrix}. \tag{2.2}$$

Here, $\epsilon_{\mathbf{r}}{}^{-1}$ is the inverse of the permittivity tensor, n is the material's refractive

index, and r_{41} is a parameter referred to as the electro-optic coefficient. Note that when no field is applied, the tensor is diagonal and is therefore isotropic, whereas the presence of a field generates off-diagonal components and induces anisotropy. Note also that the refractive index is effectively field-dependent, meaning that the EO effect is basically a glorified $\chi^{(2)}$ nonlinearity. Since this $\chi^{(2)}$ nonlinearity is measurable even at optical frequencies, this implies that the EO effect can be used to change the index of the crystal extremely quickly, on timescales of just a few femtoseconds. Detecting terahertz fields (on picosecond timescales) is no problem for the EO effect.

The basic idea of EO detection is to detect polarization rotation induced in the probe pulse. This can be done quite easily with the so-called balanced detection scheme illustrated in Figure 2-1. If the EO crystal is oriented with its [110] direction along the direction of beam propagation, then fluctuations in the probe's polarization can be detected using a quarter wave-plate, a polarizing beamsplitter, and a pair of photodiodes. The Jones matrix associated with the crystal-waveplate system is given by

$$M = \begin{pmatrix} \frac{1+i}{2} - \delta & \frac{1-i}{2} + \delta \\ \frac{1-i}{2} + \delta & \frac{1+i}{2} - \delta \end{pmatrix}, \tag{2.3}$$

where δ is the phase shift of the probe light inside the EO crystal induced by a vertically-polarized electric field. Using elementary operations it can be found to be $\delta = \frac{1}{2}\frac{\omega_{\mathrm{NIR}}}{c} n^3 E_0 r_{41} d$, where d is the thickness of the EO crystal and E_0 is the applied electric field. If the probe pulse is initially vertically-polarized with an intensity I_0, then the intensity will be split nearly evenly into each polarization, with the vertical polarization receiving an intensity of $I_V \approx (\frac{1}{2} - \delta) I_0$ and the horizontal polarization receiving a intensity of $I_H \approx (\frac{1}{2} + \delta) I_0$. By measuring the difference in power between each polarization, one finds that the difference between them is $2\delta I_0$, or

$$\frac{\Delta I}{I_0} = \frac{\omega_{\mathrm{NIR}}}{c} n^3 E_0 r_{41} d \tag{2.4}$$

It is this balanced detection measurement that represents the terahertz field. As with the photoconductive switch measurement, the final result is field- and not power-

sensitive, permitting the phase of the terahertz pulse to be measured.

One must consider several things when designing an EO detection system. The first is the choice of EO material. By looking at Equation 2.4 one might assume that the optimal crystal is a thick crystal with a high r_{41}. In this regard, ZnTe should outperform GaP since r_{41}(ZnTe)=4 pm/V and r_{41}(GaP)=1 pm/V [57]. Nevertheless, this assumption ignores the effects of the mismatch between the phase velocity of terahertz light in the crystal and the group velocity of the probe pulse. One can conceptualize the two fields as co-propagating in the crystal, with the terahertz continuously rotating the polarization of the probe pulse as they interact. But because the final measurement is field-sensitive, if the probe pulse walks sufficiently far off the terahertz field its polarization rotation will actually reverse directions, reducing the measured signal. This has two major consequences. The first is that crystals cannot be made arbitrarily thick without reducing their generation bandwidth. The second is that materials whose infrared group velocities match their terahertz phase velocities over the frequencies of interest will have better performance. Even though GaP has a lower r_{41} than ZnTe, its generation bandwidth is usually better than ZnTe's since ZnTe has a phonon resonance at 5.3 THz that greatly reduces its group velocity above a frequency of about 3 THz. Figure 2-5 illustrates this effect dramatically by showing the same TDS spectrum obtained with different crystals. Even though the ZnTe crystal has a larger peak SNR than the GaP crystal, the bandwidth of the GaP crystal is more than 6 THz, while the ZnTe can barely achieve 3 THz.

Finally, note that as balanced detection ultimately boils down to a measurement of the optical phase of the probe pulse, its fundamental noise limit is the probe shot noise [60]. (This is a manifestation of the well-known energy-time uncertainty principle.) Even though amplitude noise of the mode-locked laser is typically much higher, the use of balanced detection removes most of this noise and allows for shot noise-limited detection. Therefore, the noise floor of such systems is given by

$$I_{\mathrm{RMS}} = \sqrt{2e\eta I_0}, \tag{2.5}$$

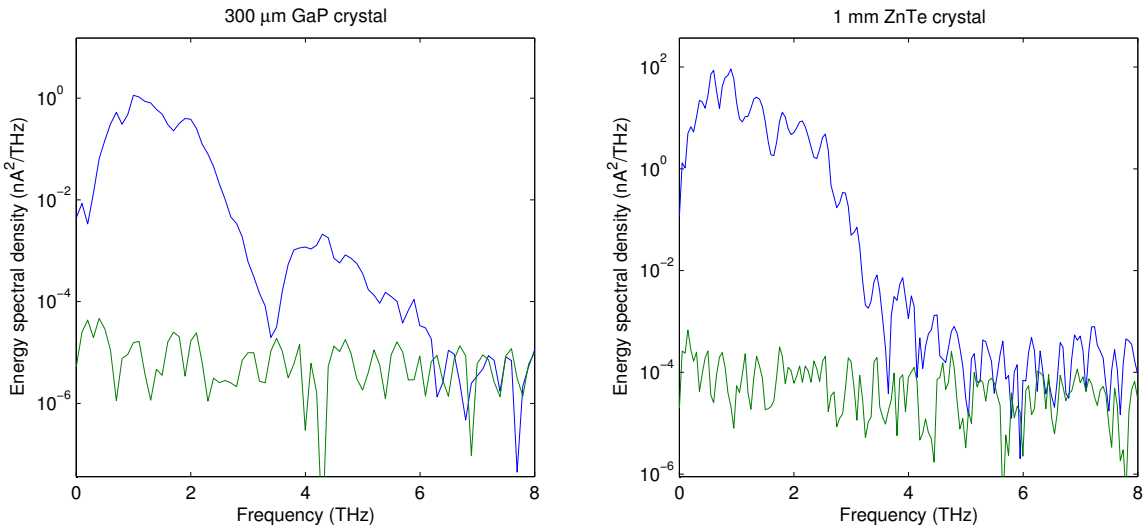


Figure 2-5: Terahertz detected by a 300 µm GaP crystal and a 1 mm ZnTe crystal under the same pump conditions, along with the corresponding noise floors. (The dip at 3 THz is due to transmission through a QCL.)

where e is the electron charge and η is the detector efficiency ($\eta \sim 0.5$ for Si). For a typical probe power of 10 mW, this leads to a noise floor on the order of 0.04 nA/$\sqrt{\text{Hz}}$. Even though the noise goes up with $\sqrt{I_0}$, the signal goes up faster, with I_0. This means that turning up the probe power is always beneficial provided that the detectors are not saturated.

2.1.3 Frequency coverage and dynamic range

Because THz-TDS records the Fourier transform and electric field waveform instead of the power spectrum directly, its resolution and dynamic range will be determined by the effective windowing of the time-domain signal. Any mechanical delay element has a finite travel distance, and this will affect the minimum achievable resolution of the system. From basic Fourier theory, the frequency resolution of a stage with travel distance L will be $c/(2L)$, which results in a resolution of 15 GHz for a stage with 1 cm of travel. However, there are usually practical limitations that limit it much further. The first is signal-to-noise ratio. When the signal being measured has no high-resolution features in the frequency domain, needlessly measuring them by

increasing the scan length reduces the SNR that can be achieved in a certain time interval. This is because all of the pulse's energy is concentrated within a certain time interval, and measuring the waveform outside that interval adds no signal but contributes noise.

The second major restriction on resolution is created by reflections in the system. Any interface present in the beam path that causes either the terahertz or the near-infrared light to split and return later will create satellite pulses, which leads to frequency-domain fringes and reduces the achievable resolution. Though almost every interface can be avoided by judicious choice of optics, there is a strong one that cannot be avoided almost by definition: the EO sampling crystal itself. When terahertz light exits the crystal it reflects and returns later, creating small pulses after the main one. When probe light exits the crystal and reflects, it samples the terahertz pulse later, making it appear as if a piece of the terahertz pulse had arrived early. This effect is shown in Figure 2-6a, which shows that a 300 µm GaP crystal creates a secondary reflection that occurs $2n_{\mathrm{GaP}}L/c$=8 ps after the initial one. (This particular pulse has passed through a QCL, and also has a second pulse that occurs much later as a result of intracavity reflection.) One solution to this problem is to use a thicker crystal so that satellite pulses are pushed further out, but as discussed previously this will reduce the terahertz bandwidth that the crystal is able to sample. A better solution is to bond the sampling crystal to a slab of electro-optically inert crystal made from the same material, e.g. a [100]-oriented crystal. Though the reflection is inevitable, wafer bonding delays it substantially and greatly increases the resolution that can be achieved. For example, Figure 2-6b shows the same pulse with a bonded crystal. The secondary reflection from the GaP has been pushed out to 40 ps and is not even visible over this time range.

The dynamic range of TDS measurements is affected by several factors. Naturally, the noise floor of the measurement and the peak signal level impose one limit on the dynamic range. The other major limitation is the shape of the window function that is applied to the measured pulse in the time domain. Though the finite travel of the stage suggests that a boxcar function is the maximum-resolution window, this window

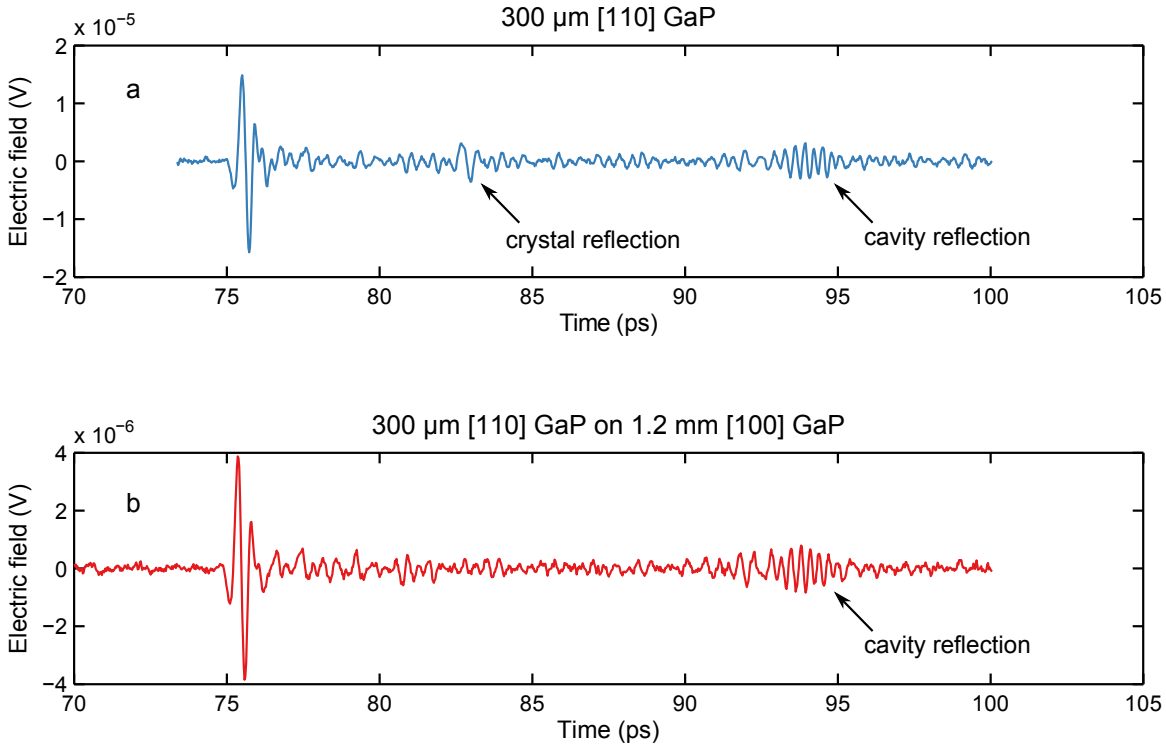


Figure 2-6: Long-travel terahertz pulses obtained from a QCL, sampled with (a) a 300 µm [110] GaP crystal, and (b) a 300 µm [110] GaP crystal wafer-bonded to 1.2 mm of [100] GaP.

has a low dynamic range, especially when the choice of limits creates a discontinuity at the edges. Typically, a more moderate window function is used, principally cosine-like Hamming, Hanning, or Blackman windows. Since these windows taper off at the edges, it is important that the pulse of interest be centered in the center of the window to prevent the signal of interest from being overly-apodized.

2.2 Time domain studies of THz QCLs using THz-TDS

Because THz-TDS is so useful for probing samples, it is also extremely useful for characterizing THz QCL gain media, on account of its unique ability to identify both the amplitude and phase associated with intersubband transitions. Since these types of measurements were first demonstrated by Kroll et al. [39], other researchers

have used the technique to demonstrate the presence of linewidth narrowing [61], and spatial hole burning [40]. Typically, THz radiation is generated externally with a photoconductive antenna, focused onto one QCL facet, transmitted through the waveguide, collected from the other facet, and detected through photoconductive or electro-optic means. Because this process leads to low coupling efficiencies, only QCLs with single plasmon waveguide geometries—which have large modes and a Fresnel-like impedance mismatch to free space—were previously characterized with this method. However, metal-metal waveguides have largely supplanted the surface plasmon waveguide in THz QCL research, as their tighter confinement of the optical mode leads to a near-unity overlap with the active region and reduced mirror losses [23], resulting in much higher maximum operating temperatures [28, 30]. On the other hand, these very properties make coupling efficiencies to and from the waveguide with an external THz pulse emitter low, thereby making TDS difficult to perform.

Several schemes were proposed to overcome the small coupling efficiency of THz radiation into metal-metal waveguides, including the integration of horn antennas [62] and the affixation of hyperhemispherical lenses to the QCL facet [51]. One of the most promising methods demonstrated earlier was the generation of terahertz pulses at the facet itself via photoconductive means. In this scheme, femtosecond near-infrared pulses were shined directly onto the facet of an operating laser. Terahertz pulses are generated by the resulting current transient [63, 64]. However, since the pulse amplitude (and possibly its frequency dependence) depends on the QCL bias, on-facet intracavity generation fundamentally couples the generation process of the THz pulse to the amplification process of the QCL, making it difficult to determine how the pulse is modified by the active region. By the same token, the excitation affects the QCL: free electrons generated by NIR excitation have been shown to increase the loss of the waveguide [63], and voltage transients generated by instantaneous shorting can affect the QCL bias.[65]

2.2.1 Photoconductive antennas integrated into QCLs

Rather than use the laser itself as photoconductive switch, a better two-section scheme was used. Essentially, a separate photoconductive emitter section was fabricated from the QCL ridge itself, with the emitter and laser ridge separated by an air gap on the order of a few microns wide [66]. Because the emitter is by definition well-matched to the QCL, the generated pulse couples to the QCL with high efficiency. Light essentially "tunnels" from one waveguide to the next. Since the emitter can be independently biased, an unambiguous determination of the active region's effect on the generated pulse can be made.

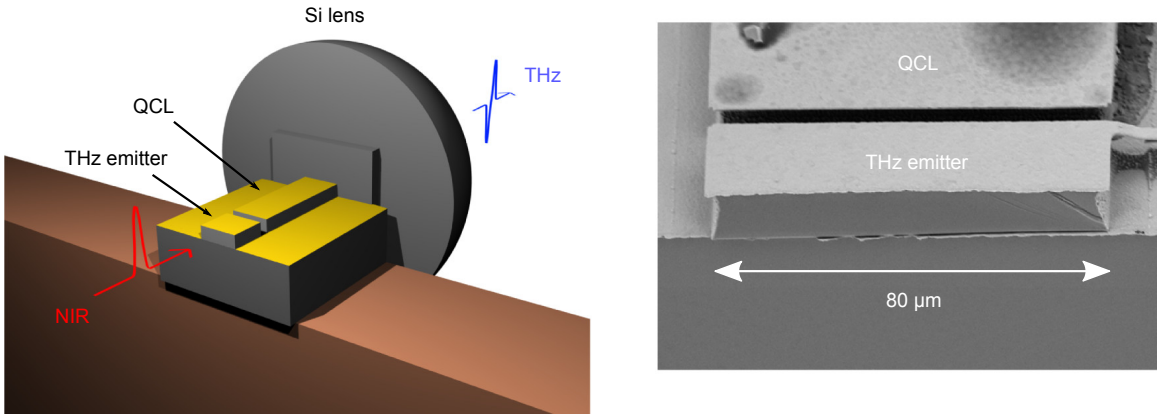


Figure 2-7: Schematic of integrated emitter used to generate terahertz pulses and QCL.

Figure 2-7 shows a basic diagram of typical photoconductive switches that were fabricated from the laser gain medium for the initial batch of devices. The gap that separates the emitter section from the laser section was defined lithographically with the same plasma etch that creates the QCL ridge, and was 4 µm wide. The emitter section was fabricated to be 100 µm long, but was subsequently cleaved to 15 µm in order to reduce the effects of Fabry-Perot fringes within the emitter. Not shown on the inset is the bonding pad used to bias the emitter, which was connected to the emitter with a lithographically-formed 4 µm gold strip and was fabricated by masking a QCL ridge with an insulating silicon dioxide layer before depositing the

top metal contact. Aplanatic silicon hyperhemispherical lenses were usually affixed to the back facet of the QCL in order to improve out-coupling of the terahertz pulse, as described in Ref. [51]. The use of this lens, together with a proper alignment, is critical for obtaining high signal-to-noise ratios, but comes at the cost of slightly higher waveguide loss.

2.2.2 Cleaved two-section devices

Though QCLs with integrated emitters were initially used as photoconductive antennas, it was ultimately determined that non-integrated emitters could perform even better in some ways than their integrated counterparts [67]. These devices were formed by micro-manipulating two QCLs with cleaved facets together; a schematic of these devices is shown in Figure 2-8. Note that if the two pieces come from different wafers their substrates may have differing thicknesses.

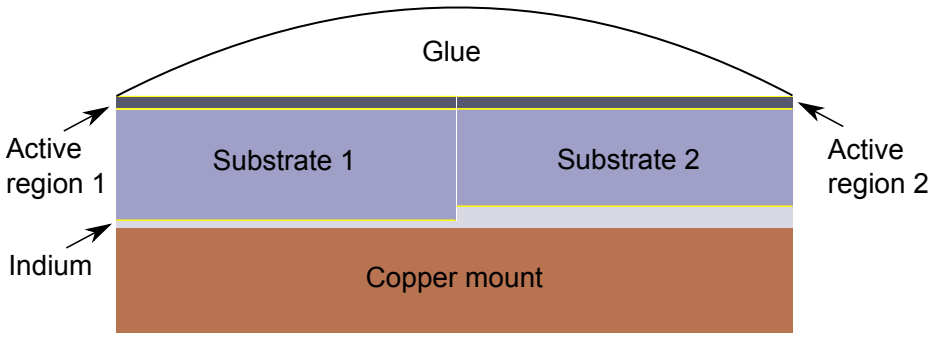


Figure 2-8: Schematic of two-section devices constructed by cleaving. The glue is eventually stripped.

This technique has several advantages. The first is that it greatly increased the quality of the interface between the two sections. Because the reactive ion etch required a deep etch (10 µm), the smallest gaps that could be achieved reliably were about 4 µm. Moreover, the quality of the sidewalls that could be achieved with this sort of etch was originally very poor. (Figure 2-9 shows gaps that were typical of the RIE process.) The result of this meant that the coupling efficiency between the two sections was quite low, estimated from its Fabry-Perot reflections to be about 50%. In

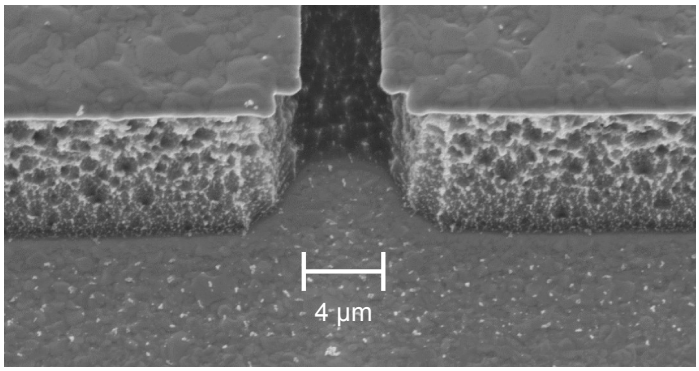


Figure 2-9: SEM image highlighting the shortcomings of RIE for making gaps in two-section devices.

contrast, cleaved facets are essentially perfectly flat—see Figure 2-7—meaning that there was no scattering loss between the two sections. In addition, provided that the two sections were pushed together and aligned properly, the gap size could be made sub-micron. The second advantage of this technique (and perhaps the most helpful) was that it could be done with lasers that were already fabricated, obviating the need for a month-long processing run. It also meant that testing could be performed on lasers that had been fabricated using wet-etching techniques, and on devices that were not suitable as photoconductive emitters. To make these devices, the following process was used:

1. The QCL emitter section and the QCL laser section were cleaved and micro-manipulated together. Alignment was verified using microscopes placed above and to the side of the gap.

2. The two lasers were cemented together using UV-curing Norland 81 optical glue, placed on top of the lasers.

3. The lasers were indium-soldered to a copper mount at 155°C. Thermal expansion of the optical glue ensured that the two sections would separate by a small amount, typically 0.5 µm. This small but finite gap ensured that the lasers remained electrically isolated.

4. The optical glue was stripped from the lasers using acetone.

This process could also be simplified for devices whose emitter and laser were made from the same chip. Instead of micromanipulating two individual chips together, one chip would be glued from the top and subsequently cleaved from the bottom. This ensured that the two pieces remained in close proximity to each other, replacing steps 1 and 2. This process is illustrated in Figure 2-10.

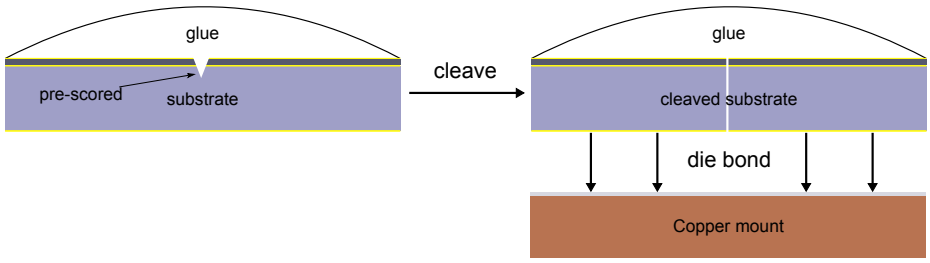


Figure 2-10: Simpler two-section device fabrication technique, in which devices are glued and subsequently cleaved.

2.2.3 Other experimental details

All of the time-domain measurements were performed in a closed-cycle pulsed-tube cryocooler, equipped with a Cryomech PT60 cryorefrigerator. The cooler is able to maintain devices at a temperature of 30 K indefinitely when no heating is applied, and has a cooling power of about 1.2 K/W. This is especially advantageous for time-domain measurements, since a full sweep of laser bias and temperature requires approximately a week of continuous measurement. The device mount used for these measurements is equipped with a temperature sensor and a heater that can be used to control the device temperature by sinking up to 50 W. This would not typically be enough power to keep the lasers at temperatures above 100 K, but this could be accomplished by simply artificially reduced the cooling power of the cold head by placing a sheet of plastic between the mount and cold head.

A simplified schematic of the beam path used for most of these measurements is shown in Figure 2-11. Light from the Ti:Sapphire (a Spectra-Physics Tsunami) propagated into a set of four mirrors designed to provide external pulse compression (labeled as DCMs). It was then split, and the pump path was attenuated using a

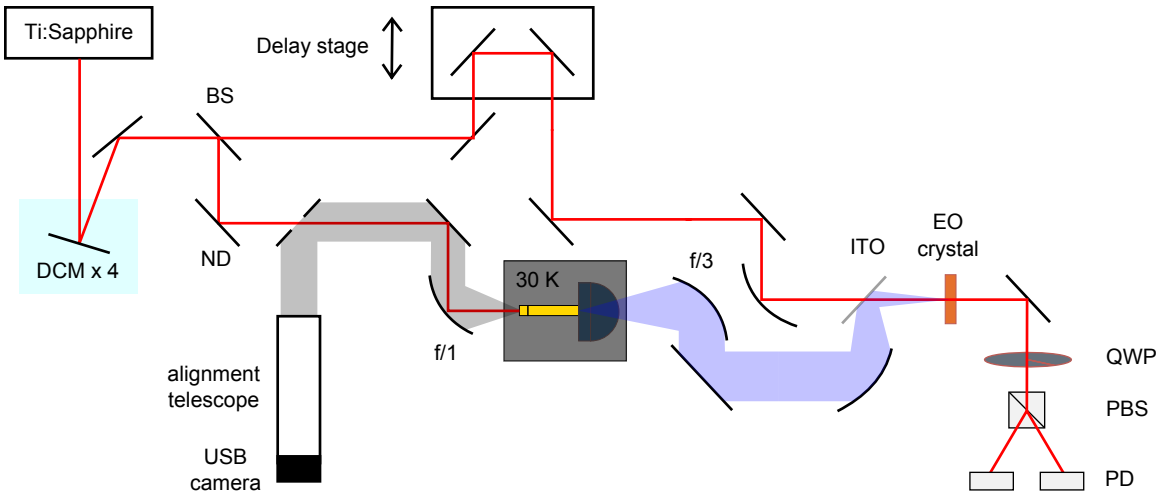


Figure 2-11: Optical path used for most time-domain measurements.

reflective neutral density filter to provide anywhere from 25 mW to 125 mW of pump power. Note that preserving the short duration of the optical pulses required that dispersive optical elements needed to be avoided, as their effects would be to unnecessarily broaden the optical pulses. This precluded the use of lenses and necessitated the use of off-axis parabolic mirrors (OAPs) for most of the focusing optics. The pump light was shined onto the QCL emitter using an f/1 mirror.

To ensure accurate alignment of the pump beam onto the QCL, an alignment telescope equipped with a USB camera was built, and together with the f/1 mirror they functioned as a microscope with which the QCL and the spot could be simultaneously viewed. Figure 2-12 shows one such view. While using this microscope to view the device, it quickly became apparent that the cold head of the cryostat was not mechanically stable, and that the position of the device varied in both the short term—fluctuating at 2.4 Hz, the compressor's operating frequency—and also in the long term—contracting by several millimeters when it was cooled and whenever the temperature was adjusted. Therefore, mechanical feedback was implemented in the form of stepper actuators placed on the objective mirror. A software-based PID controller kept the device in the center of the alignment microscope image, effectively moving the beam in concert with the device. This feedback was critical for

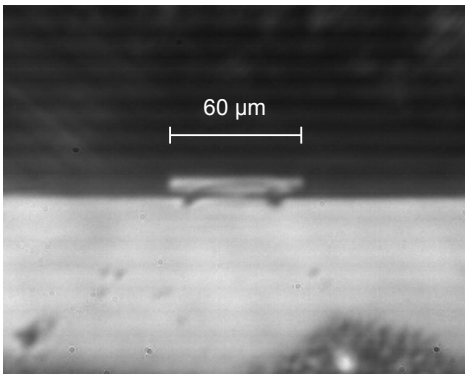


Figure 2-12: Image of QCL as viewed through alignment microscope.

experiments lasting more than a few hours, including long temperature series.

The terahertz light that came out of the QCL was collected with an f/3 mirror, and was refocused and combined with the probe light by means of an indium tin oxide (ITO) dichroic mirror. As ITO is conductive at low frequencies it also reflects very effectively in the terahertz while simultaneously transmitting visible and near-infrared. (This property is what makes it useful as a transparent electrode, for example in displays.) The two beams were made to copropagate through the electro-optic crystal, and EO sampling was used to detect the terahertz light as a function of mechanical delay.

2.2.4 Spectroscopic reference

One difficulty that arises from the use of photoconductive emitters inside double metal waveguides is the issue of obtaining a clean spectroscopic reference. Measuring the transfer function of a system requires that the signal be measured both through the sample and without the sample. However, even though integrated emitters allow much of the terahertz light to be collected, they do not offer a way of obtaining a clean reference. The main issue is that the fabrication and experimental procedures are simply not repeatable enough to allow for measurements of multiple devices to be very consistent. This means that self-referencing must be done: that is, both the signal and the reference must be measured with one device alone.

Various self-referencing schemes were attempted, including collection of the backward-generated terahertz and of the terahertz that circulated within the cavity multiple times. However, these did not prove to be reliable, especially when dealing with deep absorption features that required a high dynamic range. The only scheme that produced very reliable results was a so-called double modulation scheme, in which the emitter and the QCL were modulated at different frequencies and the resulting signals were processed using multiple lock-in amplifiers. This technique allows the user to determine what the terahertz signal is with the QCL on and with the QCL off. There are several ways of accomplishing this, the details of which are provided in Appendix B.

The most important consequence of this kind of reference is that absolute transmission of a terahertz pulse through the waveguide cannot be determined: it will always be measured relative to the device's state when it is off. As a result, gain measurements will often have bias-independent features that look like gain, but are in actuality absorption features at low bias that have merely been reduced at higher bias levels.

2.2.5 Basic results

The first QCL with a metal-metal waveguide that was successfully characterized using THz-TDS was a four-well resonant phonon design that lases at around 2.2 THz (design FL175M-M3, wafer EA1222). The frequency is relatively low for THz QCLs, but was amenable to TDS since photoconductive sources are more effective at low frequencies. Integrated emitters were used, as shown in Figure 2-7, and were excited using light from with a duration of 80 fs, a repetition rate of 80 MHz, and a center wavelength of 790 nm. These parameters were found to maximize the amplitude of the generated terahertz pulse. The 125 mW pump pulse was then focused onto the emitter facet to a spot size of about 20 µm. The choice of such a large spot size compared to the QCL height (10 µm) helped to reduce the effect of mechanical oscillations induced by the pulsed-tube cryostat, which cooled the QCL to 33 K. Radiation was collected out of the device and focused onto a 1-mm thick ZnTe crystal with a pair of f/2.5 off-

axis parabolic mirrors, and standard electro-optic sampling was performed to detect the electric field as a function of probe delay. As in Ref. [64], the peak field scaled approximately linearly in emitter bias, and though a reverse-biased emitter produced a peak field similar in magnitude to a forward-biased emitter, it also minimized dark current. This aspect of the reverse bias, together with the modest choice of pump power, allowed similar emitters to be biased as high as -50 V before devices failed.

For this measurement, the emitter was square-wave modulated at 50 kHz from -16 V to +1 V, and the QCL was independently modulated at 27 kHz with 10% duty cycle. The balanced detector output was fed into a lock-in amplifier (PAR Model 5209) tuned to the 50 kHz emitter signal, and the demodulated but unfiltered output of this amplifier was fed into a second lock-in tuned to the 27 kHz QCL signal. Asynchronous double modulation (see Appendix B) was used to determine what the transmitted pulse would be if the QCL were completely on and what it would be if it were completely off. Though this could also be done in principle by adding the difference signal to a separate reference scan taken with the QCL simply off [68], this method compensated for the aforementioned long-term drift of the QCL position in the cryostat. When the emitters were cleaved to be shorter, the terahertz spectrum obtained with the QCL off was similar to the unamplified spectra in Ref. [63], with a signal-to-noise ratio of 50 dB at 2.2 THz and a useful bandwidth of 3 THz (limited by the pump laser and the ZnTe sampling crystal).

Figure 2-13 shows an example result of this measurement for a device which was 1.21 mm long and 80 µm wide, biased above laser threshold. Both signals were effectively acquired simultaneously; though $E_{\mathrm{on}}(\tau)$ shows oscillations resulting from gain at the lasing frequency while $E_{\mathrm{off}}(\tau)$ does not. The time window shown here was only 22 ps, a typical window used for analyzing amplitude and phase data, but one can also extend it much further to see the effect of the laser cavity on pulse generation. This is shown in Figure 2-14, which shows the difference in terahertz fields obtained with the laser on and with it off. It also shows the corresponding power spectrum and a zoomed-in view of a portion of it. (Since the delay stage used had only 25 mm of travel corresponding to 166 ps of delay, five separate scans were stitched together.)

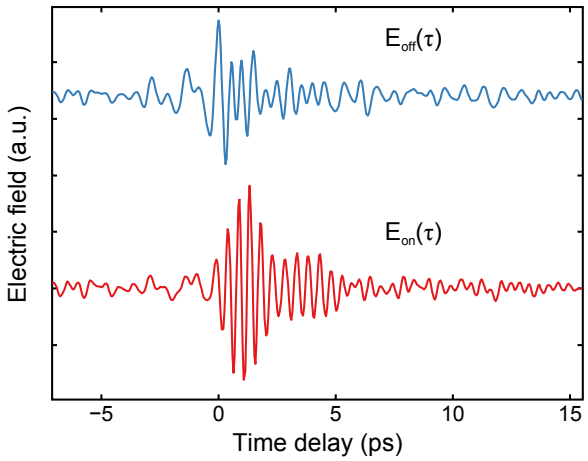


Figure 2-13: Comparison of pulse detected with QCL off and with QCL on (biased above threshold at 403 A/cm^2).

Since the signal at the lasing frequency should in principle suffer zero loss while that at nearby frequencies does not, the initial pulse generated by the emitter gives rise to dozens of echoes that correspond to the round-trip time of the cavity, 31 ps, and narrow to the lasing frequency as they circulate in the cavity. From this round-trip time, the waveguide group index $n_g \equiv c/v_g = n + \omega\frac{\partial n}{\partial \omega}$ can be inferred to be 3.84. Since the power spectral density of the difference pulse shows a uniformly-spaced comb, this indicates that only one lateral mode was excited by the THz pulse. This is advantageous from an analysis perspective, because it helps to ensure that individual echoes can be windowed for gain calculations. In addition, note that the pulse will continue to decay as it circulates through the cavity, rather than growing without bound. This is because the laser gain is clamped in such a way that the round-trip gain is zero. Each cavity round trip effectively applies a Gaussian filter to the initial broadband pulse, reducing its energy to a vanishingly small bandwidth. If one turns the laser gain on after the terahertz pulse has arrived, as was eventually done in Ref. [69], then what happens instead is that the terahertz pulse seeds the laser field and essentially becomes the lasing field.

Lastly, note the presence of a small pulse that appears 11.2 ps before the main pulse, which occurs as a result of a mode with a group index of 1.07. Since this group

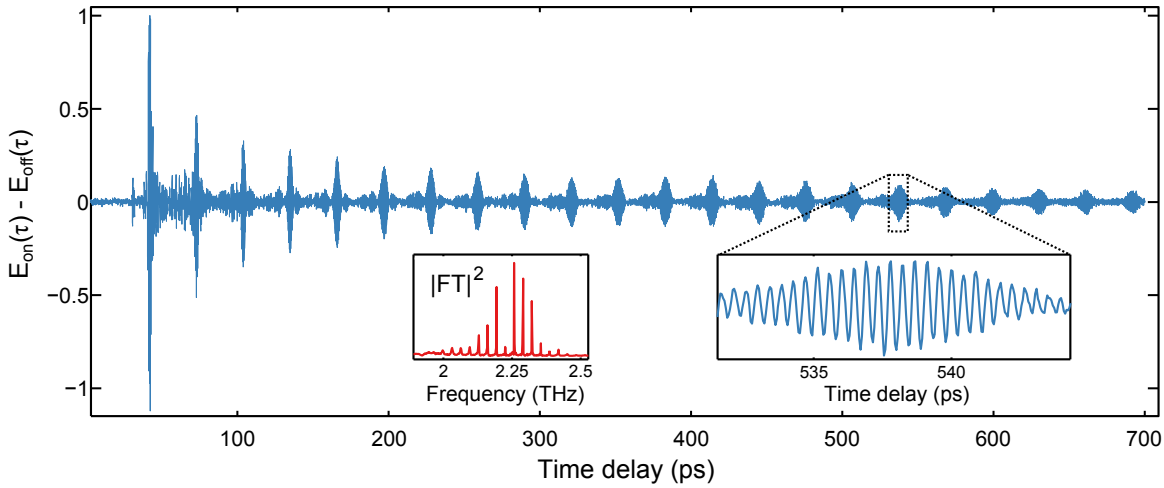


Figure 2-14: Difference in terahertz fields obtained with the laser on and off over a long travel range.

velocity is so close to the speed of light in vacuum, it indicates that it is a so-called "air-side modes." These modes were observed in electromagnetic simulations in Ref. [64] and correspond to a surface plasmon mode that propagates mostly at the corners of the top metal and in air. As they travel faster than the intracavity terahertz, they arrive first. These types of spurious modes interfere with the loss and gain analysis, but careful alignment of the aplanatic lens helps to significantly reduce this mode relative to the desired waveguide modes.

Chapter 3

Gain of terahertz quantum cascade lasers

In this chapter, the gain profile of various THz QCL design schemes are analyzed using terahertz time-domain spectroscopy. All of these designs are variations of the resonant-phonon design—the design providing the highest operating temperature of THz QCLs to date [30]. They all operate on the principle of *resonantly* injecting electrons into the upper laser state, resonantly collecting it from the lower laser state, and enticing it to emit a longitudinal-optical (LO) *phonon*. A basic schematic is shown in Figure 3-1. Modifications of the basic resonant-phonon design elicit different

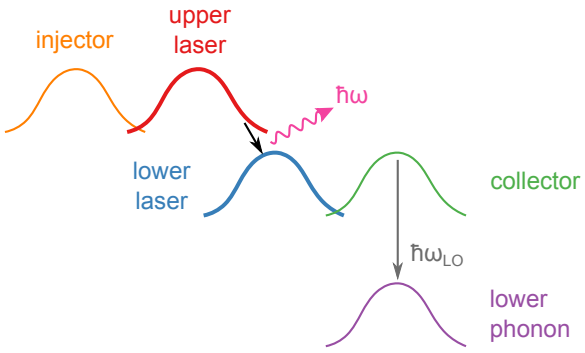


Figure 3-1: Basic resonant phonon design.

properties. The original version utilized a two-well injection scheme and was suitable for both low-frequency and high-power lasers, owing to its intrinsic stability and overall robustness. One-well injection schemes produced more interesting results—for

example, producing the current operating temperature record as well as an interesting scattering-assisted design—albeit with added cost of some unpredictability. Resonant phonon designs are probably the most interesting from the standpoint of frequency combs as well, as they tend to produce the broadest gain bandwidths.

3.1 Low frequency resonant-phonon design, FL175M-M3

This active region, previously mentioned in section 2.2.5, is a two-well injector resonant phonon design that lases near 2.2 THz. It is similar to the design described in Ref. [70] and is discussed extensively in Ref. [66]. Figure 3-2 shows its band diagram, as well as its gain spectrum measured at a few biases. Starting from the left injection

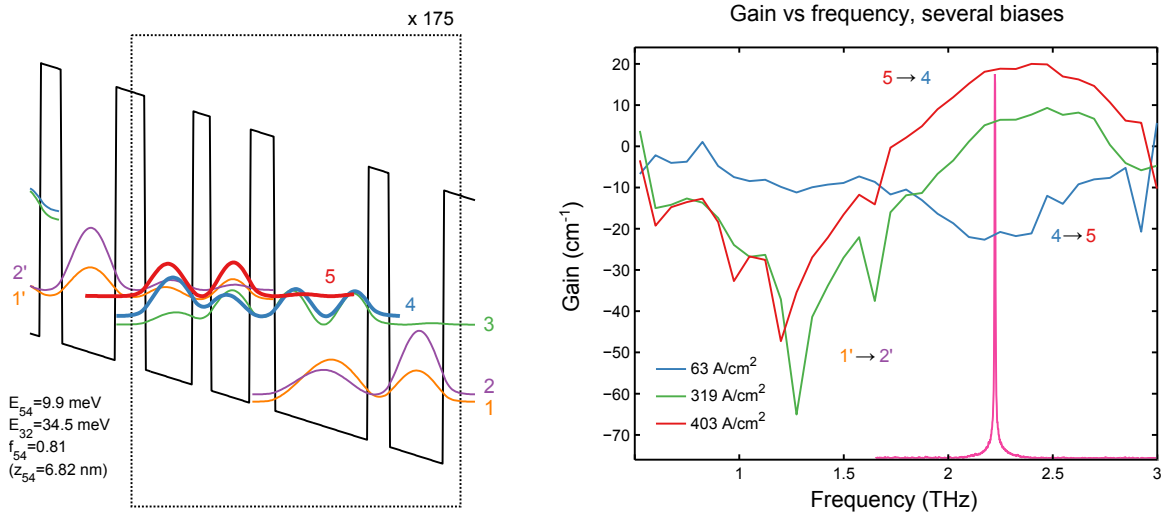


Figure 3-2: Band diagram of FL175M-M3 with some corresponding gain spectra. The lasing spectrum of a similar device is shown for reference.

barrier, the layer thicknesses of the gain medium in Å are **51**/82/**31**/68/**42**/161/**37**/93, with the bold fonts representing barriers. The 161 Å well is bulk-doped to 1.9×10^{16} cm^{-3}, resulting in a 3×10^{10} cm^{-2} sheet doping. In this diagram, state 5 is the upper laser level, state 4 is the lower laser level, state 3 is the collector state, and states 1 and 2 comprise the two-well injector.

Gain was calculated using a 13.3 ps Hamming window that excludes the air-side mode and reflections within the ZnTe sampling crystal, according to the following relation:

$$g(\omega) = \frac{1}{L} \ln \left(\frac{|E_{\text{on}}(\omega)|^2}{|E_{\text{off}}(\omega)|^2} \right) \tag{3.1}$$

Note that this expression refers to *power* gain as this is the usual convention [22], not amplitude gain as is sometimes used (e.g., in Ref. [39]). The only difference is a factor of 2, but it is important when comparing different values from the literature. The gain spectra are shown at biases of 63 A/cm^2, 319 A/cm^2, and 403 A/cm^2, corresponding to biases of far below threshold, near threshold, and above threshold, respectively. At 63 A/cm^2, far below the lasing threshold of 360 A/cm^2, the current begins to rise quickly with the bias voltage and an absorption at 2 THz is observed that can be attributed to an alignment of the injector states to the lower laser level (similar to what was seen in Ref. [65]), resulting in absorption at the lasing frequency rather than gain. A weak absorption is also observed at 1.25 THz that is likely due to absorption between the two injector states. The only other transition with a comparable energy is between levels 3 and 4, but because electrons in state 3 can quickly emit an LO phonon while the injection barrier represents a transport bottleneck, the population of state $1'$ should be significantly higher than the population of state 3. At 319 A/cm^2, still below threshold, the upper laser level is being populated, and gain is observed near 2.2 THz. The absorption between the injector states has become stronger because of a greater wavefunction overlap. Finally, at 403 A/cm^2 (above threshold), the gain at 2.2 THz has become sufficient to enable lasing, and it clamps at 18 cm^{-1} with a full-width half-maximum (FWHM) of approximately 0.67 THz. It is also apparent that the gain spectrum is asymmetric, with the lower-frequency side dragged down by the $1' \rightarrow 2'$ absorption. Clearly, this absorption in the injector region will become more severe at even lower lasing frequencies, and this is the very reason that a one-well injector scheme was developed for low-frequency resonant-phonon THz QCLs [71].

Figure 3-3a and 3-3b show the measured gain (or loss) at 2.2 THz and at 1.25 THz as a function of voltage and as a function of current, respectively, with the overlay

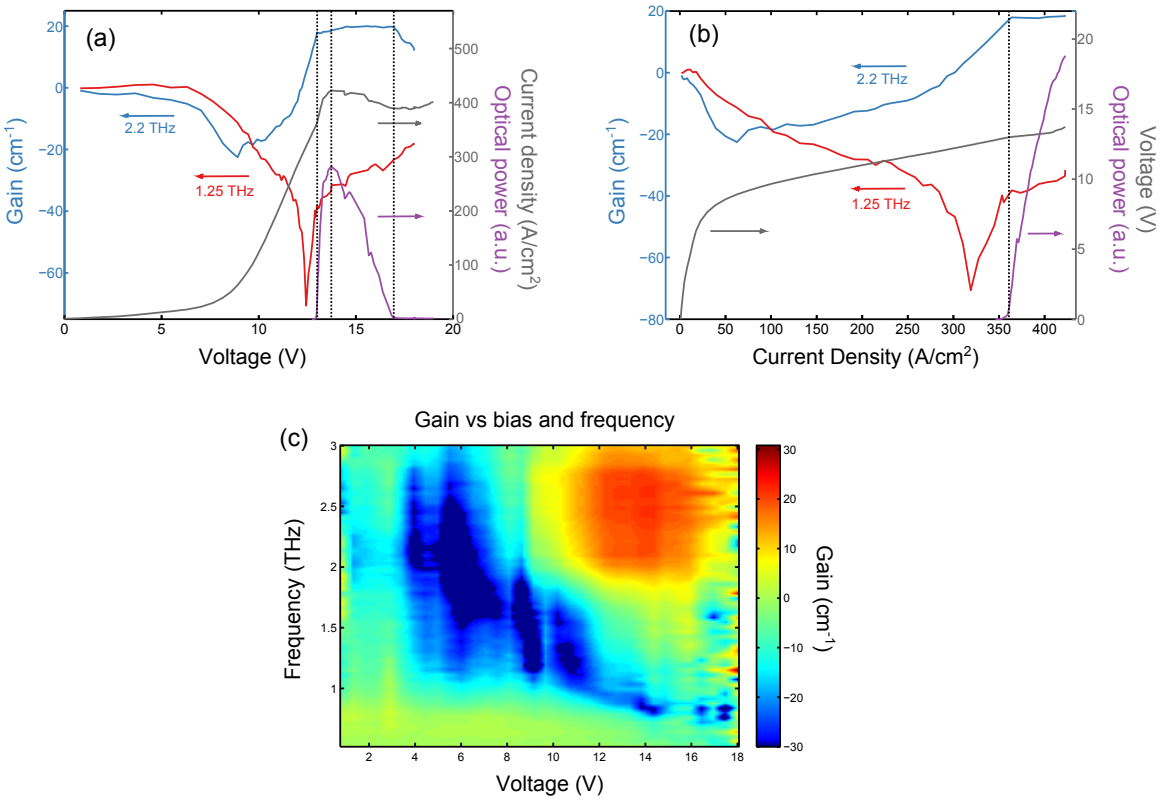


Figure 3-3: Bias dependence of FL175M-M3 gain and loss at 30 K, along with the corresponding light output.

of current-voltage (voltage-current) and power-voltage (power-current) curves shown for comparative analysis. Figure 3-3c shows the gain versus bias as a two-dimensional color plot. When the structure is unbiased, there is no absorption or gain between the lasing levels because there is no population in the lasing pair of levels 4 and 5. However, once the structure is sufficiently voltage-biased, the injector is aligned to the lower laser level 4 and absorption turns on. In fact, absorption turns on almost immediately once current begins to flow, implying that conduction within the active region at this temperature cannot occur without transport through the lasing states. As voltage is increased further, the upper laser level is populated instead of the lower one, and the absorption yields to gain. At the same time, the absorption at 1.25 THz drops, suggesting that the population inversion has come about through the depopulation of injector state $1'$: resonant tunneling is now occurring. Eventually,

laser threshold is reached, and the gain clamps to the total losses of the cavity, 18 cm^{-1}. As bias is increased to the negative differential resistance (NDR) point, the gain slightly unclamps, as electrical instability has effectively decreased the amount of time the laser is on. Finally, lasing is disabled once the structure is sufficiently biased that the gain drops below threshold.

3.2 Scattering-assisted design, OWI185E-M1

An interesting design that evolved from the resonant phonon designs was the scattering-assisted QCL [67]. The traditional means of achieving high-temperature operation relied [30, 28] on the resonant tunneling injection scheme, in which electrons are injected into the upper lasing state by resonant tunneling and leave the lower lasing state through a combination of resonant extraction and emission of LO phonons. However, this mechanism is unsuitable for high-temperature operation of QCLs operating below 2 THz, since the dynamic range of an injection-based QCL is approximately proportional to its lasing frequency. To remedy this issue, THz QCLs have been recently been developed that utilize scattering-assisted injection, in which electrons are injected into the upper state by the direct emission of an LO phonon [31, 72, 73]. In particular, the gain medium demonstrated in Ref. [31], labeled OWI185E-M1, was able to achieve lasing at 1.8 THz up to a temperature of 163 K, which is the highest operating temperature for a QCL operating below 2 THz to date. This gain medium also exhibited resonant-injection based lasing at 4.0 THz up to a temperature of 151 K, making it potentially useful in broadband heterogeneous QCLs [74, 75, 76].

In order to better understand the operation of this device, THz-TDS was used to investigate how its gain evolves with bias and temperature. Cleaved two-section emitters were used for most of these measurements, which allowed for wet-etched devices to be tested (which had superior temperature performance as a result of their lower optical losses). Note because the first section is usually quite long and is often biased below threshold, it acts as a strongly-coupled cavity and can greatly increase the total optical losses for the laser. Although this prevents a loss measurement of the

uncoupled cavity, lasing can be inhibited in such structures, allowing the measurement of unclamped gain and the properties of the gain medium alone.

The data from two lasers will be presented here: Device A was 60 µm wide and 0.6 mm long, while Device B was 40 µm wide and 1.05 mm long. Both devices were equipped with an emitter section approximately 0.5-mm long. Pulses from the Ti:Sapphire (with a 70 fs pulse width, 785 nm center wavelength, and 100 mW power) were focused onto the emitter sections, which were biased at -20 V with a 120 kHz square wave. The laser sections themselves were biased at 10% duty cycle and were triggered off of every other emitter bias pulse, so that signals obtained when the laser was on and when it was off could be compared. Electro-optic sampling was performed using a 300-µm GaP crystal, and a pair of lock-in amplifiers tuned to 60 kHz and 120 kHz were used to detect the terahertz field transmitted through the device. Synchronous double modulation was used to determine the field transmitted with the laser on and with the laser off, see B-2 for details.

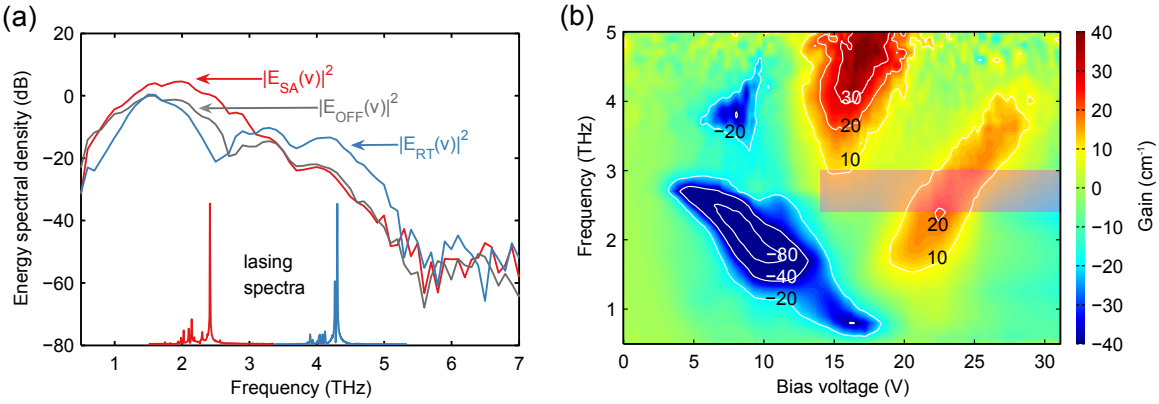


Figure 3-4: (a) TDS spectra taken from device B, along with lasing spectra. (b) Contour plot of gain at 35 K.

Figure 3-4(a) shows the frequency content associated with pulses transmitted through Device B at 35 K. It also shows the lasing spectra. When the laser is off, the transmitted pulse has a broadband frequency response, with the exception of an absorption feature at 2.7 THz. This dip corresponds to absorption from the first excited state of the unbiased module to the second, and has an oscillator strength of

about 0.7. This feature is relatively unimportant since it vanishes under design bias, but matters insofar as the gain of the QCL is always measured with respect to the gain at zero bias. Since a reduction in losses is indistinguishable from an increase in gain, all measurements at much higher biases will therefore show a "gain" that peaks at 2.7 THz, which is actually an artifact from the vanishing absorption. Fortunately, its linewidth is relatively narrow (0.5 THz), and it is located away from the peak lasing frequencies. When the device is biased to be lasing in the resonant-tunneling and scattering-assisted regimes, peaks in the transmitted spectra appear at around 4 THz and around 2 THz, respectively, corresponding to the presence of both types of gain. Note that the bandwidth of the generated terahertz pulse is quite high—above 5 THz—enabling THz-TDS measurements on most QCLs operating below the Reststrahlen band of GaAs. This is due to improvements made in the original version of the TDS system.

Figure 3-4(b) shows a contour plot of the results of a gain measurement of Device A at 35 K. Since Device A's emitter section was nearly as long as its laser section, its overall cavity losses were effectively doubled, and the device no longer lases, even at low temperatures. Therefore, Fig. 3-4(b)) is a measure of the gain medium's properties without the influence of optical feedback. The I-V characteristics of this device are not shown and are similar to the I-V characteristics of the non-lasing device described in Ref. [31]; its major distinctive feature is the lack of any obvious negative differential resistance (NDR). Finally, it should be noted that the aforementioned absorption artifact at 2.7 THz was digitally removed by interpolating between the closest frequencies unaffected by it; this region is shaded to indicate the region in which data was altered.

Figure 3-5(a) shows the gain data alongside the frequency of the relevant transitions determined by band structure simulations (whose wavefunctions are shown in Figure 3-5(c)). For ease of interpretation, two naming schemes to track their energies: low-bias wavefunctions are denoted by letters and are labeled according to their zero bias (0 V) counterparts, while high-bias wavefunctions are denoted by numbers and are labeled according to their design bias (15 V) counterparts. At low biases, there

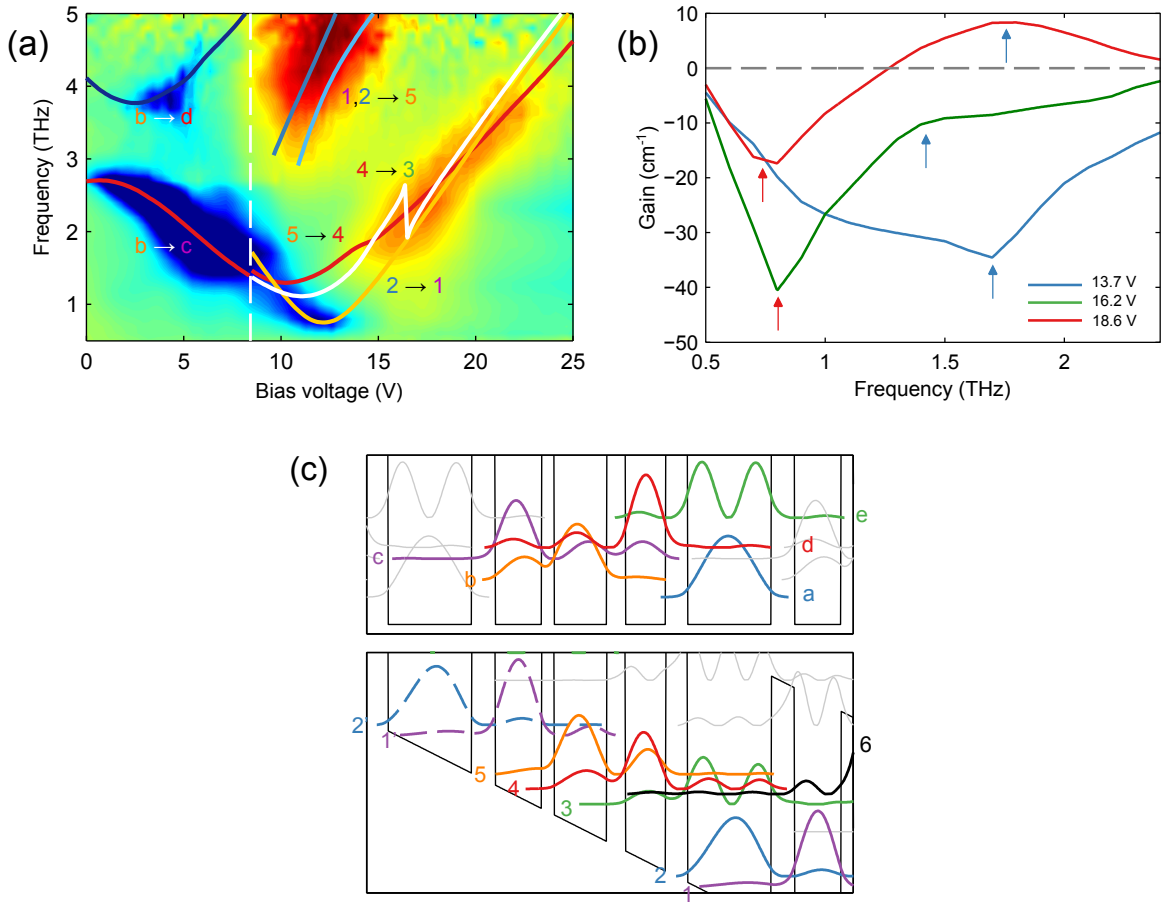


Figure 3-5: (a) Simulated transition energies superimposed on gain data. (b) Low frequency gain and absorption near design bias. (c) Band diagram of QCL at low bias and at high bias.

is a regime in which strong absorption can be observed, absorption that appears to be dominated by the b $\rightarrow$ c transition. However, a closer examination reveals that there are at least two features superimposed on each other, each of which contribute to the total absorption. This could be seen directly from the anticrossings observed in Fig. 3-5(a), but is made much clearer by Fig. 3-5(b), which shows the low-frequency characteristics of the design at several bias points in the regime in which resonant tunneling injection lasing occurs. At 13.7 V, the absorption profile is asymmetric because it is comprised of multiple superimposed peaks, but they are not resolvable. At 16.2 V, their separation is increased and they are more distinguishable, showing

one at 0.8 THz and another at 1.4 THz. By 18.6 V, the higher-frequency transition has started experiencing positive net gain, while the lower frequency transition remains lossy. Each feature red-shifts to a minimum value before blue-shifting, strongly indicating that the observed behavior corresponds to anticrossings in progress. From band structure simulations, the lower-frequency absorption can be identified as occurring between states 1 and 2, while the higher-frequency transition occurs mainly between states 5 and 4, the scattering-assisted lasing transition. It is also possible that the $4 \rightarrow 3$ transition provides some gain at this temperature, but it is expected to be weaker than the $5 \rightarrow 4$ transition thanks to its reduced upper state lifetime ($\tau_4(1 - \tau_3/\tau_{43})$ =1.41 ps, compared with $\tau_5(1 - \tau_4/\tau_{54})$ =2.70 ps) and its broader linewidth (estimated to be $\Delta\nu_{43}$=1.47 THz, compared with $\Delta\nu_{54}$ =0.98 THz).

It is valuable to compare the dynamic range and frequency coverage of the resonant-tunneling injection (RT) and scattering-assisted injection (SA) transitions, as they have some interesting differences arising from their different modes of operation. RT lasing occurs between states $1'$ and 5, peaking at about 16 V, while SA lasing occurs between states 5 and 4, peaking at about 22 V. Though the RT transition has a much higher peak gain at this temperature, its dynamic range is considerably smaller. Referring to Fig. 3-4(b), if dynamic range is defined as the range over which the gain exceeds 10 cm^{-1},the RT transition has a dynamic range of 7 V while the SA transition has a dynamic range of nearly 10 V. This is a result of the fact that the SA injection mechanism is only weakly dependent on subband alignments and can occur provided that the separation between the injector and upper lasing state is near the LO phonon separation (36 meV in GaAs). In contrast, the RT injection mechanism depends on the injector and upper lasing state being precisely aligned to achieve efficient resonant tunneling. Of course, the large dynamic range also leads the SA transition to have broad frequency coverage, experiencing substantial gain at frequencies as low as 1.5 THz (at 17 V) and as high as 4.0 THz (at 27 V). In principle, this type of gain medium could be extremely useful for applications that require frequency-agile broadband sources, such as swept-source OCT [77].

To glean more information about this design, the temperature dependence of its

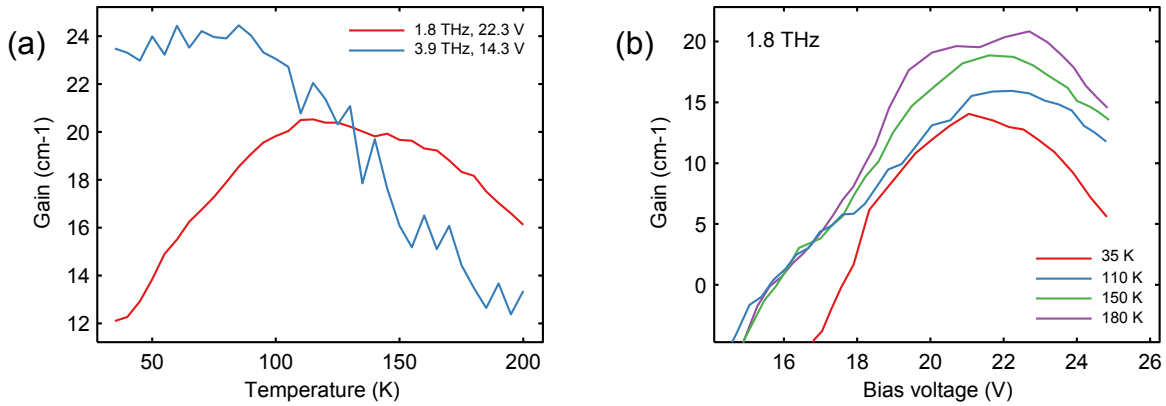


Figure 3-6: Temperature dependence of the scattering-assisted gain.

gain was analyzed. Device B was tested at temperatures ranging from 35 K to 200 K (well beyond the point at which both transitions stop lasing), and Figure 3-6(a) shows the resulting gain at the two main frequencies of interest and at fixed bias. First, consider the RT transition, shown in blue. Because the emitter section of Device B is only half the length of the laser section, this device actually lased in the RT mode at temperatures below 90 K. Therefore, gain clamping is observed at 24 cm^{-1}, a value that is approximately 50% larger than the losses expected from an uncoupled laser. This can be confirmed by considering the gain for the RT injection at Tmax=151 K, 17 cm^{-1}, which in an uncoupled cavity would be the threshold gain. The SA transition is shown in red and shows a very distinct non-monotonic behavior: as temperature is increased, its gain increases, reaching a peak value at 110 K before it begins to decay. However, this result was somewhat unexpected since the peak light output observed from a device lasing in the SA mode actually decays monotonically with increasing temperature. This discrepancy is actually due to the spectral shift of the low-frequency lasing: at low temperatures, the low-frequency lasing occurs at frequencies as high as 2.4 THz, but at high temperatures, it occurs only at 1.8 THz. Though the gain at 1.8 THz may increase, the peak optical power is determined by the gain at higher frequencies, which does decrease monotonically. The 1.8 THz gain only becomes relevant once the higher frequency components have completely

subsided; for most devices, this occurs at temperatures above 100 K.

In principle, this effect could be investigated by simply plotting the gain spectra at elevated temperatures, but in practice this is difficult since the artifact at 2.7 THz merges with the SA peak at high temperatures for frequencies above 2 THz, as the b→c transition is thermally activated and affects a larger frequency range due to its increased magnitude. It may be tempting to argue that the increase in gain with temperature at 1.8 THz is a result of similar artifact, but there are several reasons why this cannot be the case, chief of which is the fact that the measured gain at the SA T_{max} (163 K) is 19 cm^{-1}, a value which is close to the waveguide losses of 18 cm^{-1} measured in other devices with similar waveguides [66]. If flat-band absorption at 2.7 THz were significantly boosting the gain normalized to this absorption, then this value (19 cm^{-1}) would be artificially boosted. Moreover, a hallmark of such behavior is that the measured gain would not return to zero far past the design bias, but this is not the case (see Fig. 3-4(b)). In order to explain the origin of the spectral shift, plotted in Fig. 3-6(b) is the 1.8 THz gain of Device A at 35 K, 110 K, 150 K, and 180 K, swept across the SA design bias range. The peak gain at 110 K is the highest of the four curves shown, and also shows two peaks, one at 20.7 V and at 22.7 V. At higher temperatures, only one peak is clearly resolved, and the voltage at which the peak gain is observed shows a slight upward shift. Since the subband alignments in the system should be relatively independent of temperature (for constant voltage bias), this suggests that there are multiple transitions contributing to the total gain of the SA transition at elevated temperatures. Therefore, there must be an additional contribution to the gain arising from the $2' \rightarrow 1'$ transition. This contribution is of lower frequency than the 5→4 transition, and reaches 1.8 THz at a higher bias, when $E_{1'5}$ approaches 36 meV and state $1'$ is efficiently depopulated by LO phonons. This explains the dual-peaked behavior of moderate temperatures: both transitions contribute to gain at 1.8 THz, but do so at different biases. This interpretation also explains the peculiar non-monotonic behavior seen at 1.8 THz, since the depopulation of state $1'$ becomes more efficient at elevated temperatures due to increased LO phonon emission, which of course comes at the expense of gain in the RT lasing mode.

Because the linewidths at high temperatures are fairly broad, both transitions can contribute to the total lasing action near T_{max}, although the 5→4 transition becomes less effective.

3.3 Highly coherent resonant-phonon design, FL183S

A detailed look at Figure 3-5(a) will show that the resonant-phonon transition in the scattering-assisted design has been labeled as being from the result of two transitions, $1 \rightarrow 5$ and $2 \rightarrow 5$. This was no accident, and is a consequence of the fact that the two injector states, 1 and 2, are highly anticrossed and effectively form a doublet. In other words, when calculating the gain profile it doesn't make sense to talk about them in a tight-binding basis, and a delocalized basis must be used [34]. Theoretically, this should result in a gain spectrum that is split. If analyzed within the framework of density matrices, this means that the system is effectively a three-level system. An alternative framework is to consider the system as two two-level systems with phenomenological Lorentzian broadening, in which case the two upper states are separated in energy and superimpose to obtain a split gain profile.

Though this had been previously observed via the lasing spectra in various designs with large injector anticrossings, it had never been directly observed via gain measurements. Therefore, the gain spectrum of a design with a highly coherent injection scheme, FL183S, was analyzed. This design also happens to be one of the best-performing resonant-phonon QCLs in existence, producing large amounts of optical power at frequencies spanning the 3 to 4 THz range. Its bandstructure at design bias is shown in Figure 3-7. The design is a four-well design with a two-well injector, and is conceptually very similar to FL175M-M3, except it is designed for higher frequencies. Starting from the injection barrier, the layer thicknesses in Angstroms are **40**/162/**32**/92/**48**/76/**23**/73, with bold fonts representing barriers. The average doping is 5.6×10^{15} cm^{-3}. States 1 and 2 are the injector states, state 3 is the collector state, state 4 is the lower laser level, and state 5 is the upper laser level. Of note is that at design bias, the injector and upper laser level anticross by 2.4 meV, which

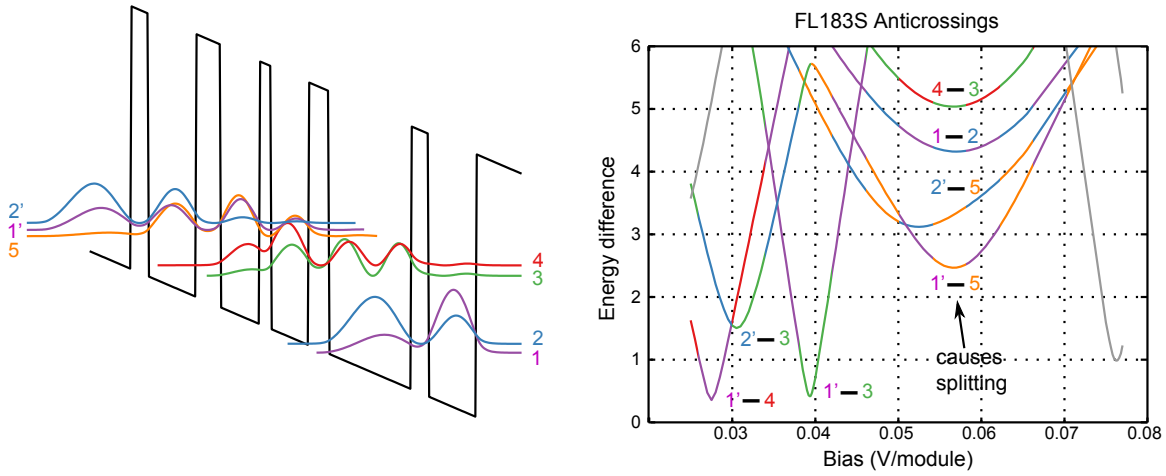


Figure 3-7: Band structure of highly coherent injection design.

should in theory lead to a splitting of the gain spectrum by 580 GHz.

However, no splitting was immediately obvious from the normal on- and off- TDS gain spectra, presumably because absorption at zero bias was causing any splitting to wash out. Therefore, an alternative self-referencing scheme had to be developed, one that utilized multiple intracavity bounces in order to provide a high-resolution profile of the gain spectrum. Figure 3-8 shows the result of this measurement, in which a single-section laser 774 µm long and 30 µm long was biased above threshold, and in which the facet was excited by 25 mW of Ti:Sapphire power. The figure shows two pulse: one that has traversed the cavity once and another that has traversed it three times. The pulses are windowed individually (using Hamming windows), with the different colors indicating the two windows. Gain was then calculated according to

$$g(\omega) = \frac{1}{2L} \ln \left(\frac{|p_3(\omega)|^2}{|p_1(\omega)|^2} \right). \tag{3.2}$$

Note that the reduction in pump power was reduced compared to earlier measurements because a single-section cavity was used. The effect of this pump was to increase the waveguide losses and cause the round-trip gain to be less than zero, and larger pumps were found to worsen this effect. The choice of a 25 mW pump proved

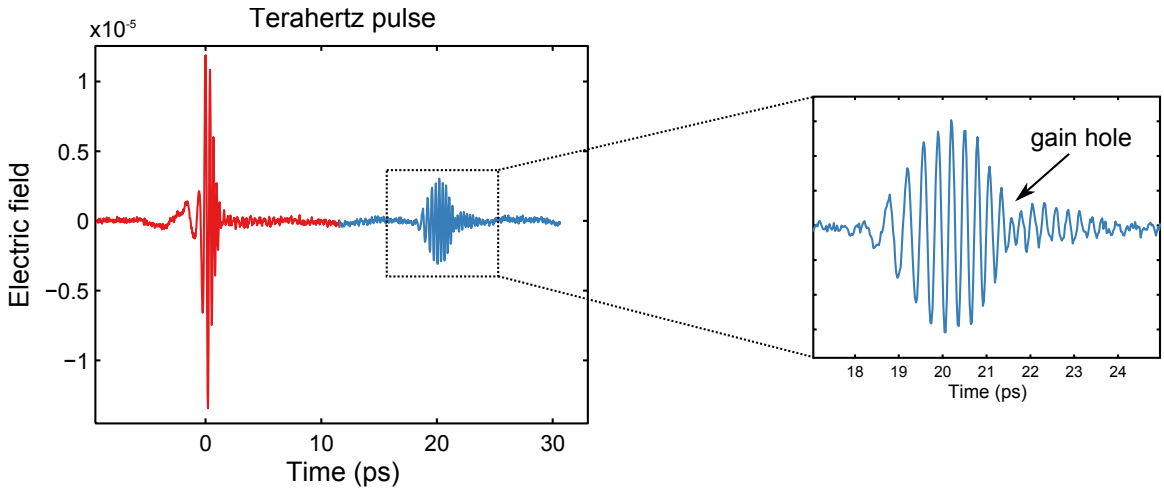


Figure 3-8: Self-referenced TDS measurement of a single-section QCL and the corresponding gain profiles.

to mitigate this effect without severely reducing the signal-to-noise ratio.

Several features are evident in the reflected pulse, which is also shown with a zoomed view. The first is that it has been substantially chirped in the time-domain. Long wavelengths are detected first, followed by short wavelengths. This is a very real representation of dispersion, and is the subject of much discussion in the following chapters. One side effect of dispersion is that time effectively becomes a measurement of frequency, which manifests very dramatically in the appearance of a "hole" in the time domain pulse at around the 21.5 ps mark. This hole effectively corresponds to a drop-off in the gain spectrum, and is a direct consequence of gain splitting.

Figure 3-9 shows the measured gain spectrum at several biases above threshold. Splitting of about 500 GHz is evident. Note that in all cases the round-trip gain peaks below zero, indicating net loss. The corresponding band stuctures are as shown, in addition to a schematic showing a tight-binding pictures of the same thing. At low biases, electrons pile up in the injector state because it is misaligned with the upper laser level. Since the injector is at lower energy, the low-frequency lobe of the gain spectrum dominates. Near design bias, the injector state and the upper laser level are maximally anticrossed, and the gain is evenly split between the two lobes. Above

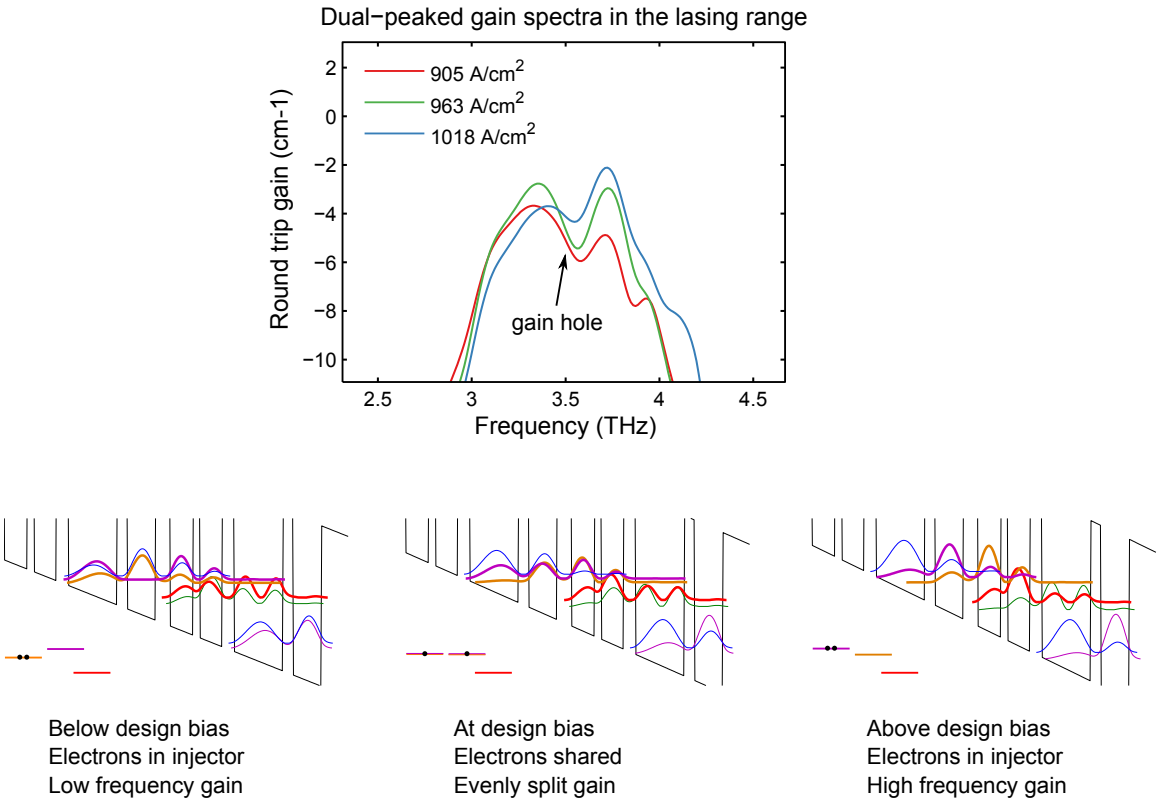


Figure 3-9: Wavefunctions of FL183S below, at, and above the design bias. Tight-binding schematics also shown.

design bias, electrons begin to pile up in the injector once again, but this time it is the higher energy configuration of the doublet and as a result the higher-frequency lobe dominates. Because the first batch of THz QCL frequency combs were based on this gain medium, the effects of this peculiar gain profile will be apparent on the comb spectra.

Chapter 4

Terahertz quantum cascade laser frequency comb design

As previously discussed, compact terahertz frequency combs would be ideal candidates for making compact spectrometers. Many ways exist for generating optical frequency combs, including intra-cavity phase modulation [78], downconversion of a higher-frequency comb [45], and upconversion of a lower-frequency source [79, 80]. However, the most powerful sources of frequency combs are generally mode-locked lasers, which in the time domain create a train of pulses and in the frequency domain create uniformly-spaced lines. Mode-locked lasers can be classified into two varieties: actively mode-locked lasers, in which pulses are generated by an external modulation of the laser gain at the round-trip frequency, and passively mode-locked lasers, in which pulses generate their own modulation with the aid of saturable absorbers. Unfortunately, because external modulation is never as short as the pulse itself, actively mode-locked lasers cannot produce pulses as short as passively mode-locked lasers. More specifically, if ω_g represents the gain-bandwidth of a laser, the best pulse width achievable by active mode-locking is proportional to $1/\sqrt{\omega_g}$, while the best pulse width achievable by passive mode-locking is proportional to $1/\omega_g$ [81]. In the frequency domain, the spectral bandwidths of actively mode-locked lasers are therefore much narrower than those of passive ones.

In quantum cascade lasers, mode-locking is difficult to achieve as the picosec-

ond gain recovery time of the laser prevents stable pulse formation [82]. Though active mode-locking of mid-infrared QCLs [83] and terahertz QCLs [84] was eventually achieved, they were unable to exploit the full gain bandwidth available to the laser. On the other hand, recent work by Hugi et al. demonstrated that when the group velocity dispersion (GVD) of a mid-IR QCL is made sufficiently low by using a broadband heterogeneous gain medium, such devices can passively form frequency combs based on four-wave mixing (FWM) [85]. Using this approach, the authors increased the bandwidth of mid-IR QCL combs from the 15 cm^{-1} achieved using active mode-locking to over 60 cm^{-1} (at 7 µm). This mechanism is in principle very similar to microresonator-based frequency combs, which were originally formed by pumping microresonators with high-intensity radiation [86]. (These combs are also called microcombs or Kerr combs.) Though this mechanism does not produce time-domain pulses with high peak intensities, the frequency comb produced is suitable for linear applications such as dual-comb spectroscopy [87, 88] and optical coherence tomography [77]. This chapter describes some of the general principles, and also how such combs can be generated in THz QCLs.

4.1 Principles of microcomb formation

Figure 4-1 shows a schematic of how microcombs are formed. A high-Q resonator is pumped with narrowband radiation at one of its cavity modes. Initially, degenerate four-wave mixing (created by the $\chi^{(3)}$ nonlinearity) causes the pump to split into sidebands that are on either side of the pump wavelength. After this has occurred, non-degenerate four-wave mixing allows for more frequencies to be generated, provided that the energy is conserved. This energy conservation is what ensures that the frequencies are all evenly-spaced and that a comb is formed. Mathematically, this can be treated simply using an envelope equation-type formalism such as that used in Boyd [92]. First, an equation that describes evolution of an EM wave in a general nonlinear wave is developed. Then, this relation is applied to the case of four-wave mixing.

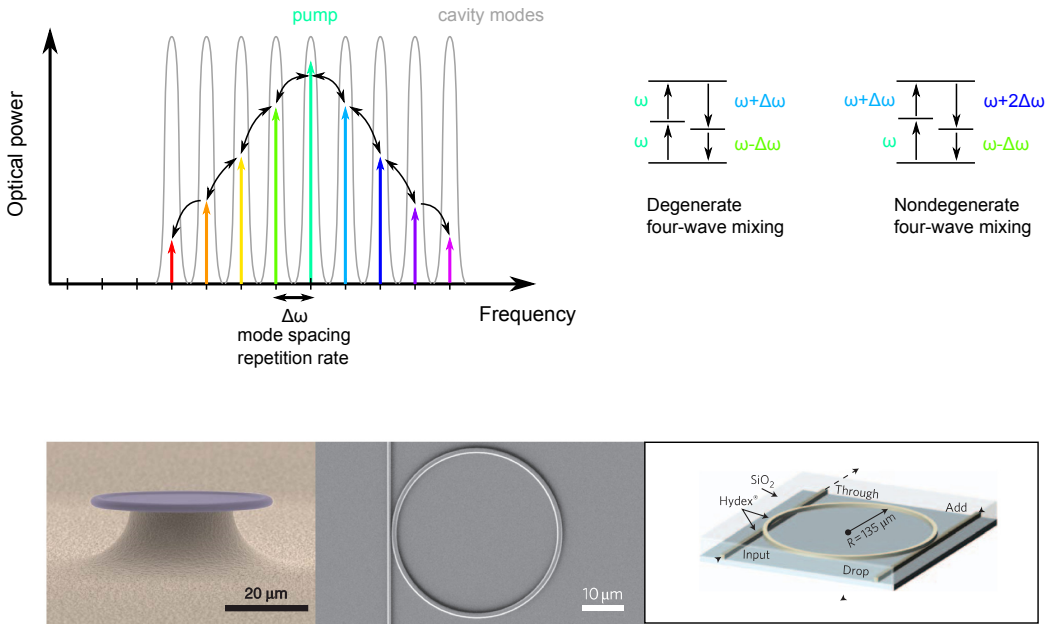


Figure 4-1: How four-wave mixing generates microcombs. Some platforms in which microcombs have been developed: silica microtoroids [89], silicon nitride rings [90], silica rings [91].

4.1.1 Nonlinear wave equation and envelopes

Maxwell's equations in a source-free region are

$$\nabla \cdot \mathbf{D} = 0 \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$$

To deal with nonlinear field evolution, the constitutive relations $\mathbf{B} = \mu_0 \mathbf{H}$ and $\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$ are assumed. Here $\mathbf{P}$ is the polarizability of the material. Under the usual assumption of plane waves (which have $\nabla \cdot \mathbf{E} = 0$), one can then derive the following wave equation:

$$\nabla^2 \mathbf{E} = \frac{1}{\epsilon_0 c^2} \frac{\partial^2 \mathbf{D}}{\partial t^2}$$

Next, the polarizability is assumed to be comprised of a linear part and a nonlinear part; that is $\mathbf{P} \equiv \epsilon_0 \chi^{(1)} \mathbf{E} + \mathbf{P}_{\mathrm{NL}}$. (Generally the linear part will be much larger than

the nonlinear components.) With this assumption, the constitutive relation becomes

$$\mathbf{D} = \epsilon_0(1+\chi^{(1)})\mathbf{E} + \mathbf{P}_{\mathrm{NL}} = \epsilon_0 n^2 \mathbf{E} + \mathbf{P}_{\mathrm{NL}}$$

where n is the usual linear refractive index of the material. Inserting this into the wave equation, one finds that

$$\boxed{\nabla^2\mathbf{E} - \left(\frac{n}{c}\right)^2 \frac{\partial^2 \mathbf{E}}{\partial t^2} = \frac{1}{\epsilon_0 c^2}\frac{\partial^2 \mathbf{P}_{\mathrm{NL}}}{\partial t^2}} \tag{4.1}$$

This modified version of the classical wave equation of course reduces to it in the absence of non-linearity, and is well-suited for treating all sorts of nonlinear processes (e.g., second harmonic generation, difference frequency generation, four-wave mixing, etc.).

Lastly, note that unless very short optical pulses are being considered, the nonlinear evolution of fields will be slow, and a good approach is to use an envelope formalism, in which one assumes that a field propagating in the z-direction with a frequency ω and wavevector $k = \frac{\omega}{c}n$ has a spatial dependence of

$$E_i(z,t) = A_i(z)e^{i(kz-\omega t)}$$

Here $A_i(z)$ is the envelope associated with the field. The advantage of this assumption is clear when one calculates the left-hand side of the nonlinear wave equation (4.1):

$$\begin{aligned}\nabla^2\mathbf{E} - \left(\frac{n}{c}\right)^2 \frac{\partial^2 \mathbf{E}}{\partial t^2} &= \left(\frac{\partial^2 A_i}{\partial z^2} + 2ik\frac{\partial A_i}{\partial z} + (ik)^2 A_i - (i\omega)^2\left(\frac{n}{c}\right)^2 A_i\right)e^{i(kz-\omega t)} \\ &= \left(\frac{\partial^2 A_i}{\partial z^2} + 2ik\frac{\partial A_i}{\partial z}\right)e^{i(kz-\omega t)}\end{aligned}$$

Under the slowly-varying approximation $\frac{\partial^2 A_i}{\partial z^2} \ll k\frac{\partial A_i}{\partial z}$, and one can write that

$$\nabla^2\mathbf{E} - \left(\frac{n}{c}\right)^2 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 2ik\frac{\partial A_i}{\partial z}e^{i(kz-\omega t)} \tag{4.2}$$

When nonlinearity generates frequencies of ω on the right-hand side of the nonlinear

wave equation, the exponentials cancel and finding the envelope functions becomes an exercise in solving a system of coupled first-order differential equations.

4.1.2 Four-wave mixing

Four-wave mixing is generated by the $\chi^{(3)}$ nonlinearity, which is defined in the 1D case as

$$P_{\mathrm{NL}} = \epsilon_0 \chi^{(3)} E^3. \tag{4.3}$$

A general field comprised of frequencies ω_i with phasors E_i can be expressed as

$$E(t) = \sum_i E_i e^{-i\omega_i t} + c.c. \tag{4.4}$$

As shorthand, let E_+ denote the first term and E_- denote the second (its conjugate). Then the polarizability can be expressed as

$$P_{\mathrm{NL}} = \epsilon_0 \chi^{(3)} (E_+ + E_-)^3 \tag{4.5}$$

$$= \epsilon_0 \chi^{(3)} (E_+^3 + 3E_+^2 E_- + 3E_+ E_-^2 + E_-^3) \tag{4.6}$$

The E_+^3 and E_-^3 terms correspond to sum-frequency generation and are ignored. This leaves

$$P_{\mathrm{NL}} = 3\epsilon_0 \chi^{(3)} \sum_{ijk} E_i E_j E_k^* e^{-i(\omega_i + \omega_j - \omega_k)t} + c.c. \tag{4.7}$$

In other words, four-wave mixing is a process by which fields at frequencies ω_i, ω_j, and ω_k generate a nonlinear polarization at a frequency $\omega_i + \omega_j - \omega_k$. Next, the envelope formalism is used by assuming that $E_i = A_i(z) e^{ik_i z}$. Placing this in the wave equation and using the envelope approximation yields

$$\nabla^2 \mathbf{E} - \left(\frac{n}{c}\right)^2 \frac{\partial^2 \mathbf{E}}{\partial t^2} = \sum_i 2ik_i \frac{\partial A_i}{\partial z} e^{i(k_i z - \omega_i t)} + c.c.$$

for the left-hand side, and

$$\frac{1}{\epsilon_0 c^2}\frac{\partial^2 \mathbf{P}_{\mathrm{NL}}}{\partial t^2} = \frac{3\chi^{(3)}}{c^2}\frac{\partial^2}{\partial t^2}\sum_{ijk} A_i A_j A_k^* e^{i(k_i+k_j-k_k)z} e^{-i(\omega_i+\omega_j-\omega_k)t} + c.c.$$
$$= -\frac{3\chi^{(3)}}{c^2}\sum_{ijk} A_i A_j A_k^* (\omega_i+\omega_j-\omega_k)^2 e^{i(k_i+k_j-k_k)z} e^{-i(\omega_i+\omega_j-\omega_k)t} + c.c.$$

for the right hand side. Proceeding further usually requires that some assumptions be made about the fields present.

Non-degenerate four-wave mixing

First, consider non-degenerate four-wave mixing, in which case fields at frequencies ω_1, ω_2, ω_3, and ω_4 are present with $\omega_1 + \omega_2 = \omega_3 + \omega_4$ and $\omega_1 \neq \omega_2$. Consider the components of the wave equation oscillating at $e^{-i\omega_4 t}$. Only the $i = 4$ component contributes on the left-hand side, whereas on the right-hand side the $(ijk) = (1,2,3)$, $(ijk) = (2,1,3)$, and $(ijk) = (4,4,4)$ components all contribute. This results in

$$2ik_4\frac{\partial A_4}{\partial z}e^{ik_4 z} = -\frac{3\chi^{(3)}}{c^2}\omega_4^2(2A_1A_2A_3^* e^{i(k_1+k_2-k_3)z} + A_4A_4A_4^* e^{i(k_4+k_4-k_4)z})$$
$$\frac{\partial A_4}{\partial z} = i\frac{3\chi^{(3)}\omega_4^2}{2c^2k_4}(2A_1A_2A_3^* e^{i(k_1+k_2-k_3-k_4)z} + A_4|A_4|^2),$$

or more simply,

$$\boxed{\frac{\partial A_4}{\partial z} = i\frac{3\chi^{(3)}\omega_4}{2nc}(2A_1A_2A_3^* e^{i(k_1+k_2-k_3-k_4)z} + A_4|A_4|^2)}. \tag{4.8}$$

The first term in equation (4.8), the FWM term, is what allows for three frequencies to generate a fourth component even if it was not originally present. The second term is referred to as self-phase modulation (SPM), as it is independent of the other three waves and looks like an intensity-dependent refractive index.

Perhaps the most important aspect of equation (4.8) is the presence of a term oscillating with $e^{i(k_1+k_2-k_3-k_4)z}$. It states that if the total momentum is not conserved, that is if $k_1 + k_2 \neq k_3 + k_4$, then four-wave mixing will not be efficient and will

eventually be self-defeating. When the momentum is linear in ω, this term will automatically be phase-matched since energy was assumed to have been conserved. However, if $\frac{\partial^2 k}{\partial \omega^2} \neq 0$, then this is no longer the case and the generation suffers. Essentially, this is equivalent to saying that the material is dispersive.

Degenerate four-wave mixing

In degenerate four-wave mixing, three waves of frequencies ω_1, ω_2, and ω_3 are present, with $2\omega_2 = \omega_1 + \omega_3$. Consider the components of the wave equation oscillating at $e^{-i\omega_3 t}$. On the left-hand side $i = 3$ contributes, and on the right-hand side $(ijk) = (2,2,1)$ and $(ijk) = (3,3,3)$ contribute. This leaves

$$2ik_3 \frac{\partial A_3}{\partial z} e^{ik_3 z} = -\frac{3\chi^{(3)}}{c^2}\omega_3^2 (A_2 A_2 A_1^* e^{i(k_2+k_2-k_1)z} + A_3 A_3 A_3^* e^{i(k_3+k_3-k_3)z}),$$

or just

$$\frac{\partial A_3}{\partial z} = i\frac{3\chi^{(3)}\omega_3}{2nc}(A_2 A_2 A_1^* e^{i(2k_2-k_1-k_3)z} + A_3|A_3|^2). \tag{4.9}$$

Once again, there is a FWM term and a SPM term, and momentum must be conserved in order for the conversion to be efficient. Since degenerate FWM generates the first few lines of a microcomb, consider a situation in which no fields other than the pump (ω_2) are present. In the absence of phase-mismatch and SPM, degenerate four-wave mixing leads to the following coupled system for the fields A_1 and A_3:

$$\frac{\partial A_3}{\partial z} = i\frac{3\chi^{(3)}\omega_3}{2nc} A_2^2 A_1^*$$
$$\frac{\partial A_1}{\partial z} = i\frac{3\chi^{(3)}\omega_1}{2nc} A_2^2 A_3^*.$$

If the pump remains undepleted ($\frac{\partial A_2}{\partial z} = 0$), then by conjugating one of these equations, differentiating, and inserting it into the other, one can show that A_1 and A_3 experience a type of parametric gain, obeying

$$\frac{\partial^2 A_{1/3}}{\partial z^2} = \left(\frac{3\chi^{(3)}}{2nc}\sqrt{\omega_1\omega_3}|A_2|^2\right)^2 A_{1/3}. \tag{4.10}$$

As the quantity in parenthesis is strictly positive, the solution to this differential equation is evidently a growing exponential. The corresponding power gain is

$$g = \frac{3\chi^{(3)}}{nc}\sqrt{\omega_1\omega_3}|A_2|^2. \tag{4.11}$$

Even if no photons of ω_1 or ω_3 are originally present, spontaneously-emitted photons will exponentially grow until they begin depleting the pump.

4.1.3 Parametric microcombs versus laser microcombs

In traditional microcombs, a single-frequency pump initiates the comb formation and the resulting parametric gain sustains it at other frequencies. As in a laser, the gain must overcome the cavity losses for this process to occur. However, in QCLs this interpretation is not as physical, as parametric gain alone will not be sufficient for the laser to overcome its comparatively higher losses. In fact, it is the gain of the QCL that is allowing the system to oscillate, gain that is essentially provided by the imaginary part of $\chi^{(1)}$ and not by $\chi^{(3)}$. A laser with sufficient hole burning will be able to lase at a lot of different frequencies without forming a frequency comb [75], so how might FWM generate a comb in QCLs?

The answer is provided by the concept of injection locking. When energy is injected into an oscillator with an frequency close to the oscillation frequency, the system can oscillate at the injected frequency instead of the natural one. Because four-wave mixing requires energy conservation, a pair of lines separated by some spacing $\Delta\omega$ will only interact with a pair of lines of exactly the same spacing. If this process is robust enough to occur over the laser's bandwidth, a comb will form. Figure 4-2 shows how this might happen. An initially unevenly-spaced multi-mode laser begins generating sidebands via four-wave mixing, sidebands which are not in the center of the cavity modes but are close enough. (The exact condition will be discussed in the following section.) The laser injection locks off-resonance, and this process continues until either the FWM is too weak because the intensity is too low, or because the spacing is too far off. Eventually a state is reached in which many of

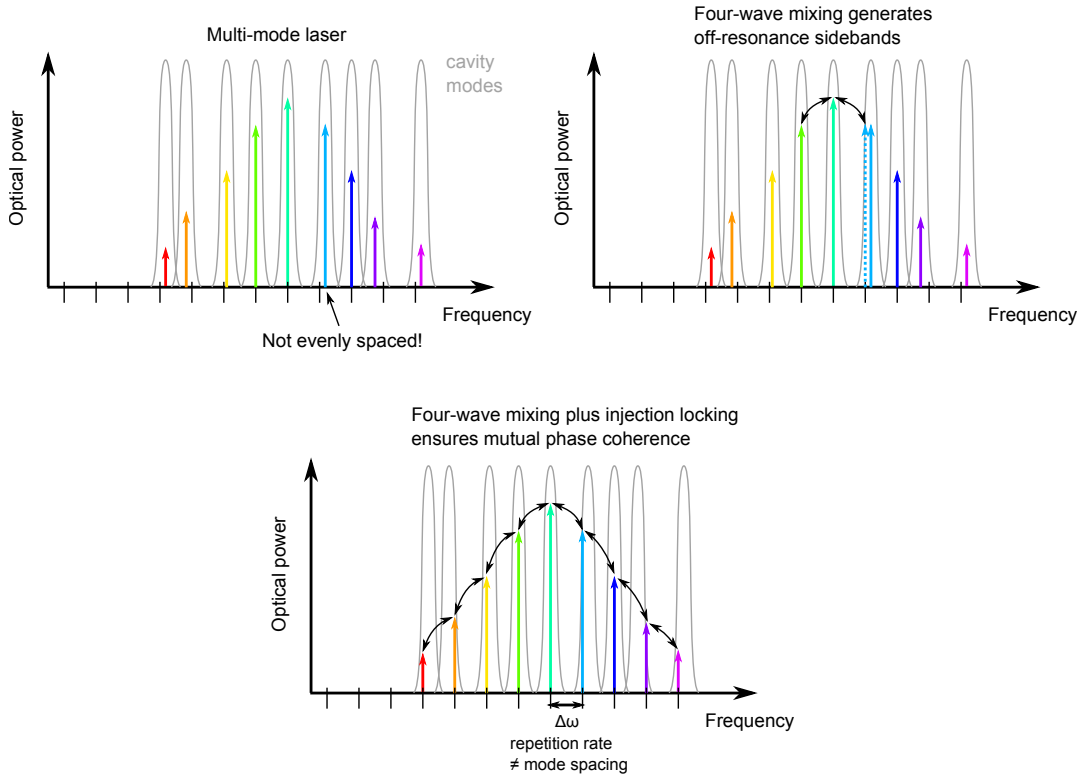


Figure 4-2: How four-wave mixing plus injection locking forms can form a comb in a highly nonlinear gain medium.

the modes have phase-locked. This process is somewhat mutually competitive, with different parts of the laser preferring different mode spacings and attempting to force this spacing on their neighbors. Only the strongest spacing survives. If a QCL gain medium has multiple lobes that prefer different mode spacings, one would expect that at different biases the mode spacing might abruptly switch depending on which lobe possessed the most optical intensity.

4.1.4 Classical injection locking

Injection locking of lasers has been around almost as long as the laser itself, and is treated in classical texts such as Siegman's [93]. Essentially, the principle (reviewed

here) is that sufficiently large injected power overcomes the losses associated with being off-resonance. Imagine that the laser cavity shown in Figure 4-3 is lasing at a frequency ω_0 with an intracavity intensity I_L, and that it has field reflectivity, gain, and loss of r, g, and α respectively. A field of intensity I and frequency $\omega \neq \omega_0$ is injected: under what conditions will the laser lase at ω instead of ω_0?

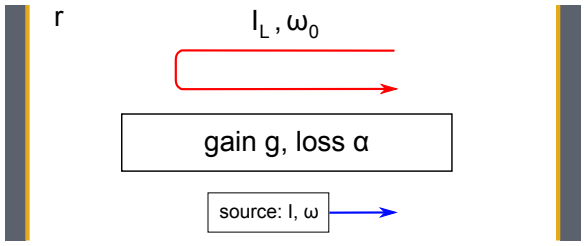


Figure 4-3: Basic laser cavity with injection locking.

Every round trip, a field inside the cavity acquires a factor of

$$\begin{aligned} A_{rt}(\omega) &= e^{i\frac{\omega}{c}n(2L)} r^2 e^{(g-\alpha)(2L)} \\ &= e^{i(\omega-\omega_0)\frac{2nL}{c}} \left[e^{i\omega_0 \frac{2nL}{c}} r^2 e^{(g-\alpha)(2L)} \right]. \end{aligned}$$

Because the system is lasing at ω_0, the quantity in brackets is known to be 1 and can be ignored. When multiple reflections are considered, the field at ω obtains a total transfer function of $A(\omega) = 1 + A_{rt}(\omega) + A_{rt}^2(\omega) + \ldots = 1/(1 - A_{rt}(\omega))$, or

$$A(\omega) = \frac{1}{1 - e^{i(\omega-\omega_0)\frac{2nL}{c}}} \approx \frac{1}{-i(\omega-\omega_0)\frac{2nL}{c}} = i\frac{c}{2nL}\frac{1}{\omega-\omega_0}. \tag{4.12}$$

In terms of the free spectral range $\Delta\omega = 2\pi c/(2nL)$,

$$A(\omega) = \frac{i}{2\pi}\frac{\Delta\omega}{\omega-\omega_0}. \tag{4.13}$$

Light that is injected sufficiently close to the lasing frequency will experience a gain of $|A(\omega)|^2$. If one imagines a weak hypothetical source of radiation at ω_0 (say spontaneous emission) with an intensity I_0, then the light at ω will dominate light at ω_0

whenever $|A(\omega)|^2 I > |A(\omega_0)|^2 I_0 = I_L$, that is whenever

$$I > I_L \left(2\pi \frac{\omega - \omega_0}{\Delta\omega} \right)^2 \tag{4.14}$$

or

$$\left| \frac{\omega - \omega_0}{\Delta\omega} \right| < \frac{1}{2\pi} \sqrt{\frac{I}{I_L}}. \tag{4.15}$$

This suggests two strategies for forming injection-locked microcombs: one can either increase the power generated by FWM or decrease the injected frequency's offset.[1] In fact, reducing the dispersion of the gain medium actually accomplishes both of these things: four-wave mixing is enhanced because better phase-matching can be achieved, and the frequency offset is decreased because dispersion is effectively the uniformity of the mode spacing.

4.2 Dispersion engineering

Clearly, dispersion is important in both traditional microcombs and laser microcombs. Unfortunately, a straightforward implementation of the approach taken in mid-IR QCLs is unlikely to scale well to THz frequencies, because terahertz electromagnetic waves strongly couple with the crystalline lattice and are thus orders of magnitude more dispersive than mid-IR waves. The coupling of EM waves to the lattice can be modeled as a simple polariton dispersion relation, which yields [94]

$$\epsilon(\omega) = \epsilon_\infty + \frac{\epsilon_\infty - \epsilon_0}{1 - \omega^2/\omega_T^2}. \tag{4.16}$$

Here, ϵ_∞ is the high-frequency dielectric constant, ϵ_0 is the low-frequency dielectric constant, and ω_T is the transverse optical phonon frequency. Figure 4-4 shows the calculated GVD of gallium arsenide (GaAs) at 40 K, using parameters obtained from Blakemore et al. [95]. Because ω_T=8 THz in GaAs, at 3.5 THz the GVD is 87,400

[1] The cold cavity Q does not appear here since laser gain compensates loss.

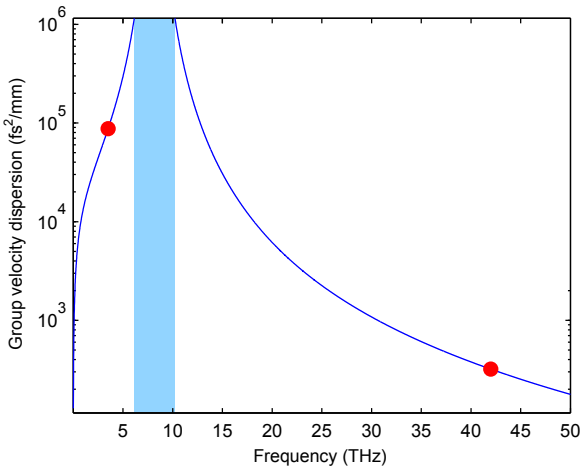


Figure 4-4: Calculated GVD of GaAs. The Restrahlen band is shown in blue.

fs^2/mm, which is 250 times greater than the GVD at 7 µm of 320 fs^2/mm. Simply using a broadband gain medium is not sufficient to overcome the dispersion of the material itself and permit comb formation.

To overcome this problem, dispersion compensation was integrated into QCL waveguides to deliberately cancel the cavity dispersion. The basic idea, illustrated in Fig. 4-5, is that a chirped distributed Bragg reflector (DBR) corrugation is etched

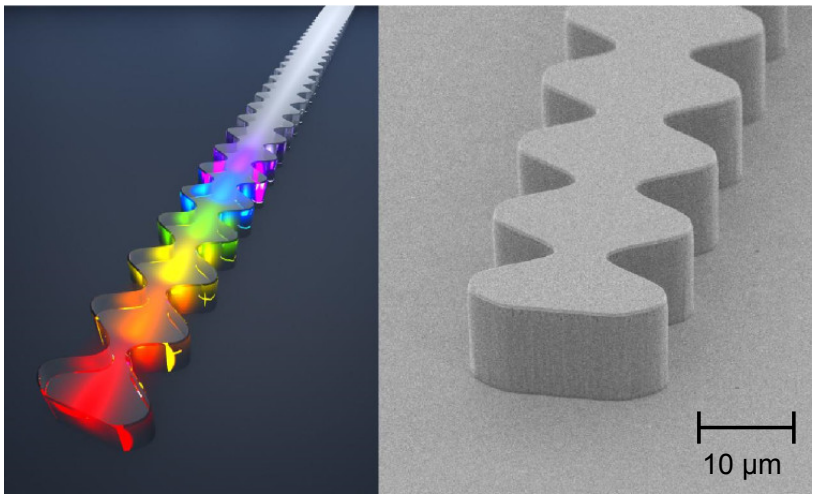


Figure 4-5: Artistic interpretation of double-chirped mirrors (DCMs) integrated into QCL waveguides, along with an SEM image.

into the facet of the laser whose period tapers from a short one to a long one as

its amplitude increases. Long-wavelength waves (which have higher group velocities) travel to the end of the cavity before reflecting, while short-wavelength waves reflect earlier, thereby compensating dispersion. Essentially, this corrugation mimics the double-chirped mirrors (DCMs) used to generate octave-spanning spectra in the near-infrared and visible range [96], but instead of being made of discrete components, the mirrors are integrated into the lasers themselves. The key advantage of this method of dispersion compensation is that it gives precise control over the group delay dispersion (GDD) of the cavity.

In order to design the corrugations used for dispersion compensation, one-dimensional simulations were first performed that captured the essential behavior of the structures by modeling the metal-metal waveguide as an infinite parallel-plate waveguide. This transfer matrix formalism is useful for capturing the essential principles of design, but is not especially accurate since it implicitly assumes a perfect magnetic conductor (PMC) at the waveguide edges. Specifically, it assumes that the impedance of a waveguide of width w, thickness t, and index n is

$$Z_i = \frac{\eta_0}{n_i}\frac{t}{w} \tag{4.17}$$

Because the model effectively over-confines the optical mode in the lateral direction, it over-predicts the extent to which the impedance of a waveguide can be increased by making it narrower. Still, it is a useful tool for guiding the more complete finite element (FEM) simulations that were eventually performed, as it can be performed in seconds instead of hours. Essentially, it functions by finding the reflectivity as a function of frequency, $\Gamma(\omega)$. Then, group delay is determined according to

$$\tau_g(\omega) = \frac{\partial}{\partial\omega}\arg(\Gamma(\omega)) \tag{4.18}$$

As dispersion is fully characterized by τ_g, this expression can be used to determine dispersion at every order. The second-order group-delay dispersion, called D_2, is

defined by

$$D_2 \equiv \frac{\partial \tau_g}{\partial \omega}. \tag{4.19}$$

Note that the units of group delay dispersion are ps^2, whereas the units of group velocity dispersion ($D_2 \equiv \frac{\partial}{\partial \omega}\frac{1}{v_g}$) is scaled by the propagation length and has units of ps^2/mm.

The key parameters, shown in Figure 4-6, are the corrugation length, its start and stop period, its stop period, how quickly its amplitude tapers on (e.g., linearly), its largest amplitude (minimum width), and end phase. Below this diagram are transfer matrix simulations showing the effect of changing these parameters on group delay. Note that in all cases the start and stop period of the DBR essentially determine the frequency range over which dispersion can be compensated. For reference, the DBR reflectivity ($|\Gamma(\omega)|$) is also shown.

The first simulation shows the effect of changing the length of the DCM while keeping everything else constant. Unsurprisingly, this effectively scales the amount of dispersion compensation (i.e., the slope of each line). This is because to first order, the corrugation length should determine the amount of compensation. The second diagram shows the effect of tapering on the amplitude with different orders. Here the amplitude as a function of position is assumed to scale in a polynomial way, i.e. $A(z) \sim z^{\alpha}$. Note that how quickly the amplitude tapers on determines how strong ripples in the final group delay are: when $\alpha = 0$ and the corrugation turns on instantaneously, large oscillations are evident. This is because the instantaneously turned-on cavity behaves like a Gires-Tournois interferometer, with the large impedance mismatch at the DCM boundary causing a substantial portion of the light to reflect immediately upon encountering the DCM. When the transmitted light eventually does reflect, it interferes with the original reflected light, creating undesirable interference fringes. While this is obviously a bad thing from the standpoint of compensating linear dispersion, it may be useful in future work if non-monotonic dispersion is to be compensated.

The third simulation shows the effect of changing the largest amplitude of the

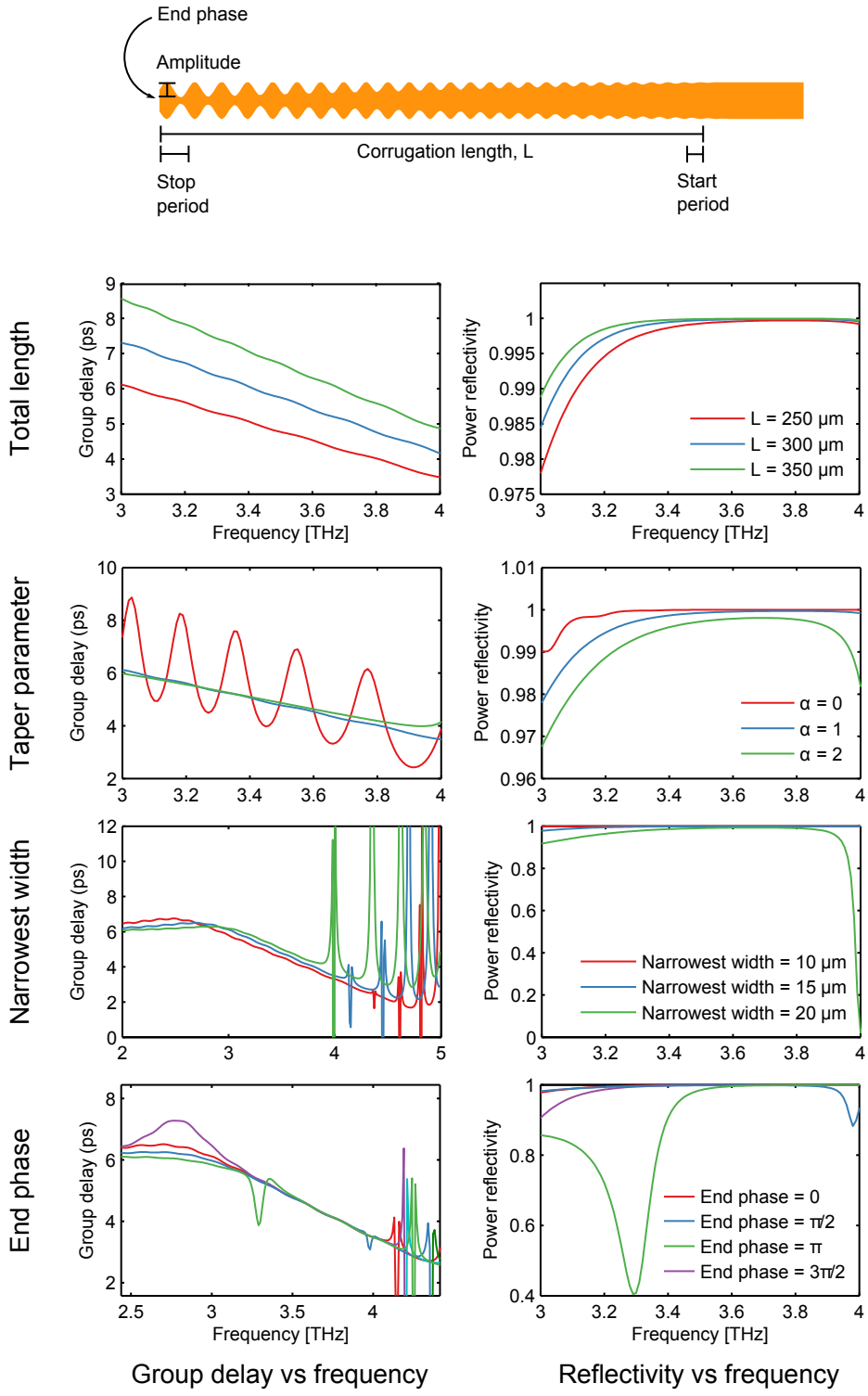


Figure 4-6: Key parameters for designing DCMs, along with transfer matrix simulations. Detailed discussion in the text.

corrugation. To a leading degree, this effect determines the bandwidth over which dispersion can effectively be compensated. When the amplitude is large (i.e., the waveguide was tapered to a narrow ridge), the range over which the dispersion remains linear increases. In contrast, when the corrugation is weaker, the range decreases. The last simulation shows the effect of changing the final phase of the sinusoid before the waveguide was terminated. (A phase of 0 is defined to be the corrugation shown, with the waveguide terminating when the ridge is halfway through shrinking. A phase of $\pi/2$ is defined as the waveguide terminating at its minimum width, etc.) Only phases of 0 were ultimately fabricated since it resulted in the cleanest GDD plot, but in the future it may be more desirable to choose a phase of $3\pi/2$ if larger bandwidths are to be compensated since it enhances the GDD at low frequencies.

Eventually, full-wave finite element (FEM) simulations were performed using COMSOL to verify design efficacy. The principle is the same as the transfer matrix simulations—group delay is found by differentiating the phase of the reflection coefficient—but the implementation is obviously different since the 3D simulations can take into account higher-order lateral modes. First, the lateral modes of the waveguide were found by performing a boundary mode analysis at the entrance of the corrugation. Since only narrow ridges 20 µm wide were fabricated, this structure could only support two lateral modes at the largest frequency of interest, 4.2 THz. Their effecive indices are plotted in Figure 4-7. Next, the S-parameters of the 3D structure were determined as a function of frequency. For a narrow ridge with two modes, symmetry and parity ensure that $S_{21} = 0$, and since a DBR has high reflectivity $|S_{11}| \approx 1$ and $|S_{22}| \approx 1$. Only the lowest order (TEM-like) mode was considered, and its reflection coefficient could therefore be taken to be S_{11}. Group delay was calculated in the same manner as before.

Figure 4-8 shows the group delay versus frequencies calculated for two well-designed DCMs, compensating dispersion of 0.215 ps^2 and 1.53 ps^2, respectively. Even though they differ in their compensation by nearly an order of magnitude, linearity is maintained over the whole design range of 3 THz to 4 THz. Though sidewall-based corrugations were used here, any perturbation that introduces a re-

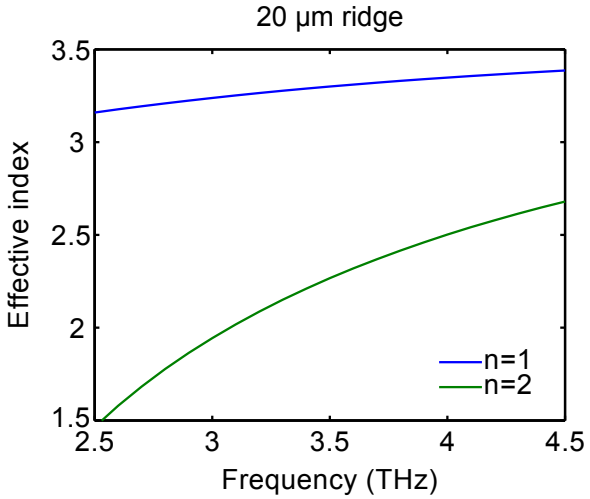


Figure 4-7: Effective mode indices for 20 μm ridge.

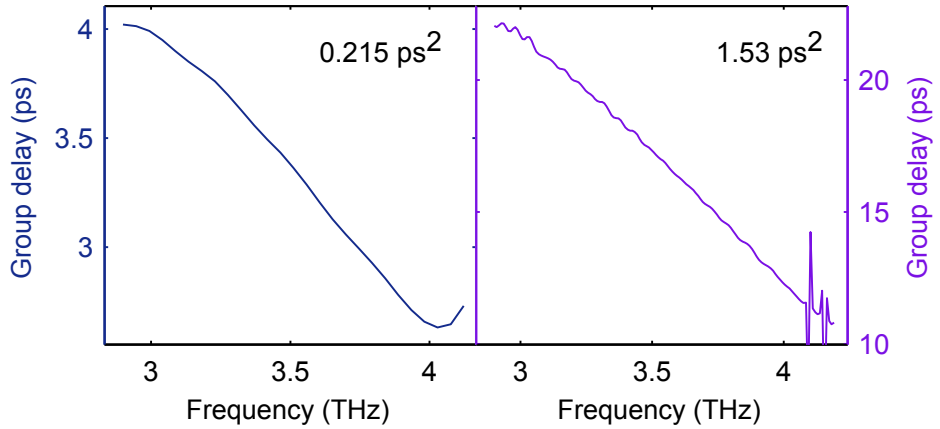


Figure 4-8: FEM simulation results for DCMs of different designs.

fractive index change into the waveguide (such as an etched trench or a region of removed metal) can also be used to construct compensators.

4.3 Actual dispersion measurement

Even though the dispersion of GaAs is known and is easily measured at low temperatures, the situation is actually somewhat worse in THz QCLs. The gain medium and waveguides also introduce dispersion into the QCL, making the dispersion unknown and possibly even frequency dependent. In order to design DCMs, an accurate

measurement of the dispersion of a real laser waveguide is essential. For this, single-section THz-TDS was performed on the FL183S gain medium, previously discussed within the context of its double-peaked gain (see section 3.3). This time, the use of a single-section waveguide allows for a cleaner measurement of the round-trip phase since the terahertz pulse does not encounter gaps that could distort the transmitted pulse. Figure 4-9 illustrates the steps involved in performing this phase measurement.

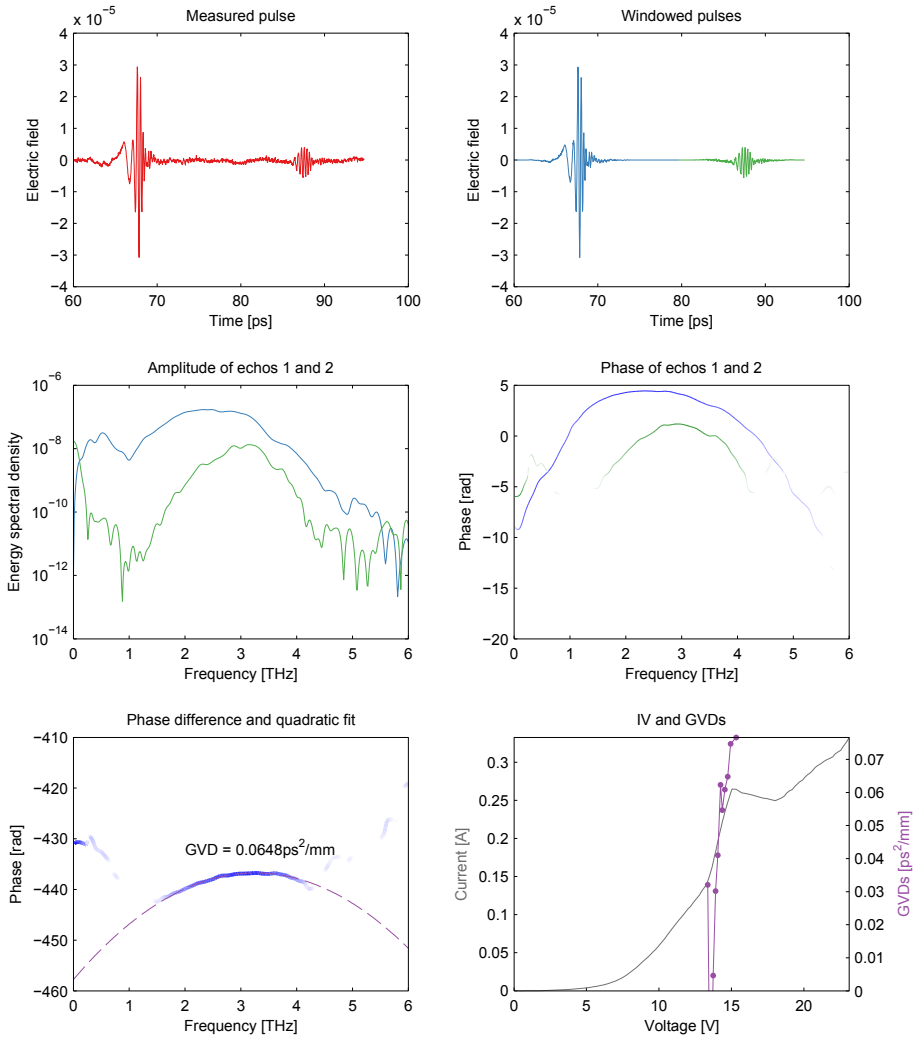


Figure 4-9: Measured GVD of a 30 µm ridge.

First, the TDS pulse, shown in red, is acquired. Next, the pulses that were measured are windowed separately using Hanning windows, and the results are shown in blue and in green. Then, the amplitudes of each pulse are plotted; the amplitude of the

blue pulse is more broadband since it has not been spectrally filtered by the laser gain. The phases are plotted by first removing their linear part (which correpsponds to trivial group delay), and then by only plotting them in regions where their SNR is large. Here, SNR is indicated by the intensity of the line. Lastly, their phase difference is plotted, and low-order polynomial fit is used to determine their total dispersion. The value of D_2 is normalized to the length of the cavity, and is plotted vs bias. Note that dispersion in the lasing regime is generally seen to be an increasing function of bias.

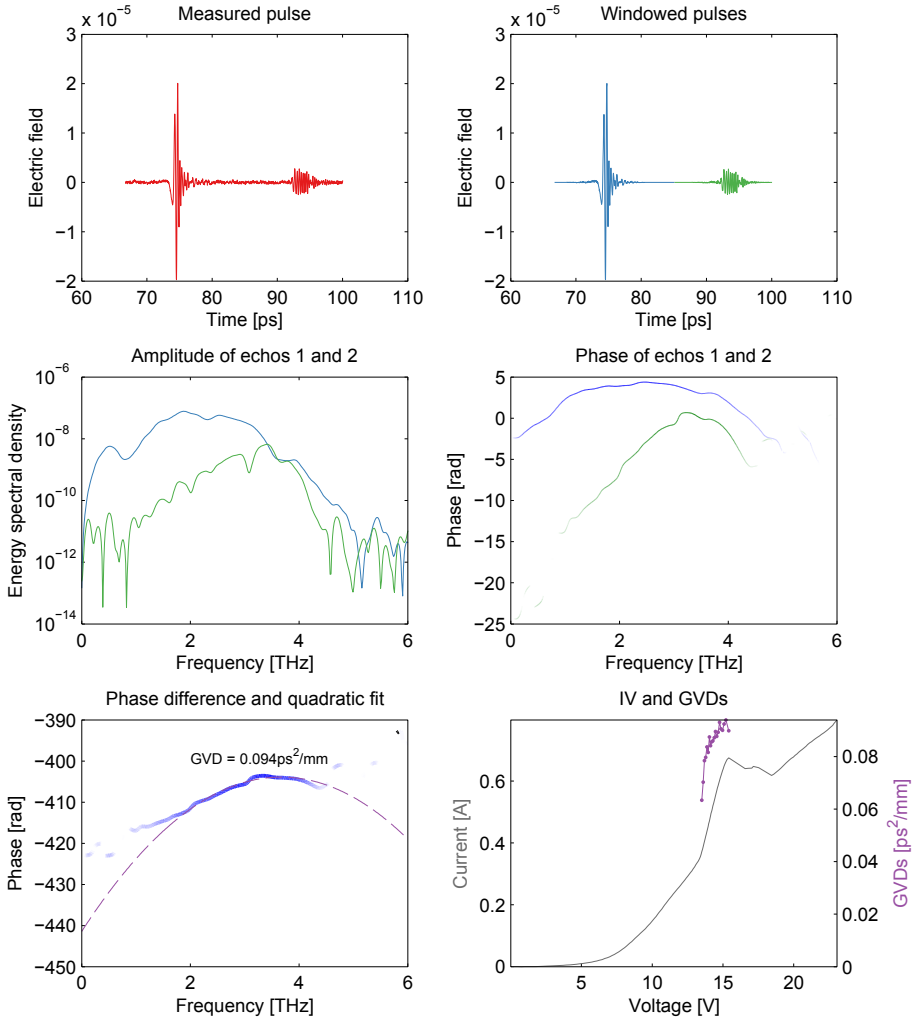


Figure 4-10: Measured GVD of an 80 µm ridge.

A key danger associated with this type of measurement is over-interpretation

of multi-mode data. Figure 4-10 shows the dispersion data measured using an 80 µm ridge instead of a 30 µm one. Even though the time-domain signal actually appears somewhat cleaner, the frequency-domain data possesses several nodes in both amplitude and phase. This is because the photoconductive generation process excites all modes which have even parity, i.e. the modes with a maximum in the center of the ridge, where the pump is shined. In this case, it is likely the n=1 and n=3 modes being excited. If two modes possess slightly different group velocities, then they will arrive at different times and produce fringing inversely proportional to their relative group delay. This effect is barely noticeable on the first pulse that exits the waveguide since the fringing is about 1 THz, but is very noticeable on the second since the fringing is three times less, about 0.33 THz.

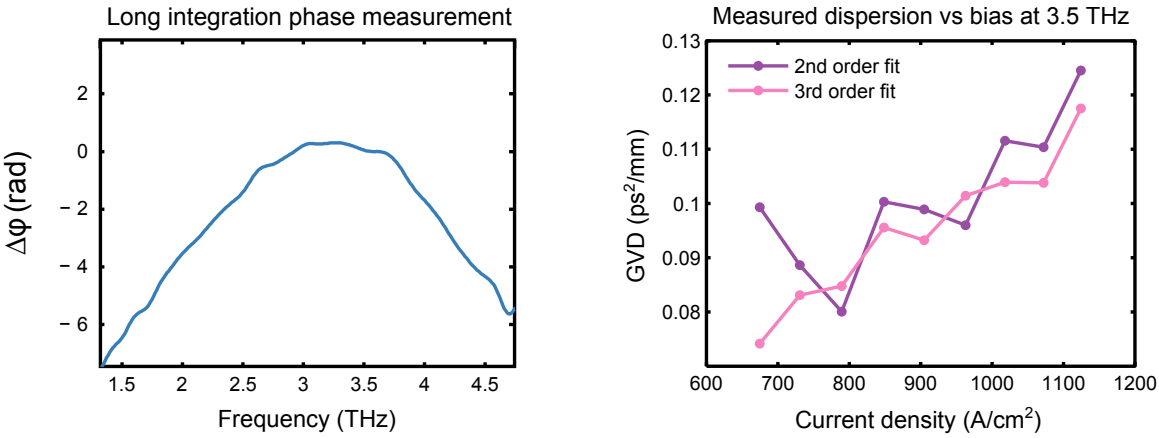


Figure 4-11: High-SNR phase measurement of a 30 µm ridge at high bias, along with the bias-dependence of measured dispersion.

Figure 4-11 shows the phase acquired by a pulse after traveling through a 775 µm long laser twice; the highly nonlinear phase-frequency relation indicates a GVD of $D_2 = 111,000$ fs^2/mm, which is substantially higher than what would be expected from GaAs alone. This value is taken to be the nominal value of D_2 at the biases of interest. The figure also shows the bias-dependence of dispersion, using both a second-order polynomial fit and a third-order fit (at 3.5 THz). In order to account for uncertainties in the measurement and in device fabrication, seven different

compensators were designed, which compensated the measured D_2 adjusted by 0%, $\pm$6.6%, $\pm$13.3%, and $\pm$20%. In subsequent measurements, only the +13.3% device was found to be interesting, whereas the rest just lased at a few modes in stable bias regimes. This coresponds to a GVD of about 125,000 fs^2/mm. The other major variation between devices was that some devices had all of their compensation on one facet (a single compensator), whereas others had their compensation split between both facets (a split compensator).

4.4 Basic results

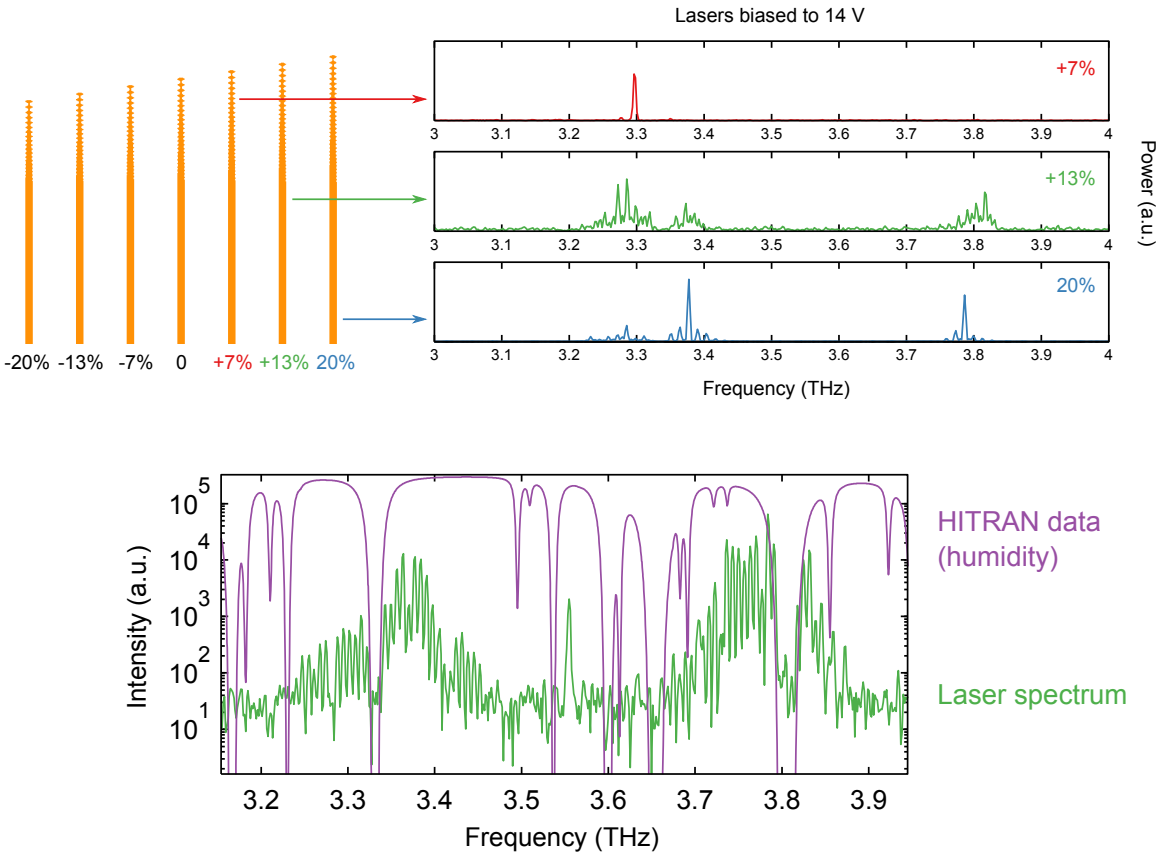


Figure 4-12: Above: DCM batch, along with the corresponding optical spectra. Below: High-resolution CW version of the 13.3% compensated spectrum.

Figure 4-12 shows the pulsed-mode spectra of a few lasers from the same batch,

with varying amounts of dispersion compensation. The lasers all have their compensations split over both facets, and the devices are all biased to 14 V at a temperature of 30 K in pulsed mode (10% duty cycle). Each laser is 5 mm long, and the spectrum was measured with a home-built Fourier Transform Spectrometer. Only the device with the 13.3% overcompensation produces an especially broad spectrum, with the +7% producing an almost single-mode output and the +20% producing a few-mode output. (More importantly, the 20% overcorrected sample does not produce a stable narrow beatnote, as will be discussed in the forthcoming chapter.) This type of behavior was seen in multiple devices of the same length, although it was eventually determined that shorter devices were more insensitive to dispersion compensation.

To further investigate the spectral properties of the 13% overcorrected devices, a hyperhemispherical lens was attached to a 5 mm device with all of its compensation on one facet. The laser is continuous-wave biased to 0.9 A, requiring that the temperature of the cryocooler be increased to 50 K to accommodate the extra heating power. The result is shown in the lower panel of 4-12, along with the calculated absorption due to atmospheric humidity. (The system was of course purged with dry nitrogen gas, but some of the spectral features are especially deep and are essentially impossible to remove without a system fully in vacuum.) As previously discussed, this gain medium has a strong injector anticrossing and possesses two peaks, one near 3.3 THz and one near 3.8 THz. These peaks are clearly visible in the optical spectra. The lines are spaced by about 6.8 GHz, roughly corresponding to the mode spacing of the cavity, although precisely determining the spacing is difficult since FTS has a limited precision. At 45 K, the lowest temperature the device could be biased at, it produced more than 5 mW of optical power, as measured with a calibrated Thomas-Keating power meter.

An interesting side effect of the broadband terahertz generation is that this same design generates a strong RF beatnote at the round-trip frequency of the laser (6.8 GHz) when DC-biased. Even without taking into account RF losses from the bond wire, as much as -33 dBm of power has been observed leaking out of the 5-mm QCLs, as measured using a bias tee. Though similar beatnotes have previously been observed

on the bias lines of metal-metal waveguides [97], at -65 dBm they were three orders of magnitude weaker. This strong RF signal implies that the beatnotes generated from all the pairs in the spectra are adding up coherently, although it is not definitive by itself. Also of note is the fact that for longer devices, only the 13.3% overcompensated devices actually produce this beatnote; see Figure 4-13 for the bias dependence of the RF power of this beatnote for several devices. Note that the "wrong" dispersion

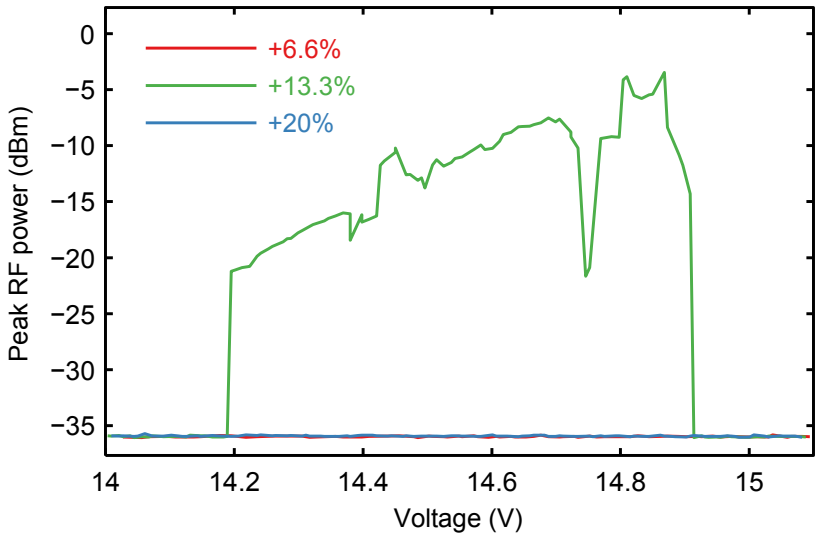


Figure 4-13: RF beatnote power measured directly from QCL using devices of varying dispersion compensation.

compensation produces no detectable beatnote. It is possible that such a beatnote is actually present, but is just too weak or too broad to be directly measured. As will be shown in the following chapter, this narrow beatnote is a signature of comb formation, as it essentially contains the optical beating between all modes in the cavity.

Figure 4-14 shows the beatnotes measured directly from 5 mm QCLs with and without lenses attached, offset relative to their carrier. The lensed devices produce beatnotes with a FWHM of about 100 kHz, while the non-lensed devices produce beatnotes whose widths are on the order of kHz. In any event, one can compare the width of these beatnotes to the expected deviation in line spacing over some bandwidth for a dispersive cavity. Suppose that two pairs of lines are separated by a frequency Δs, and that a cavity has a length L. The difference in round trip group de-

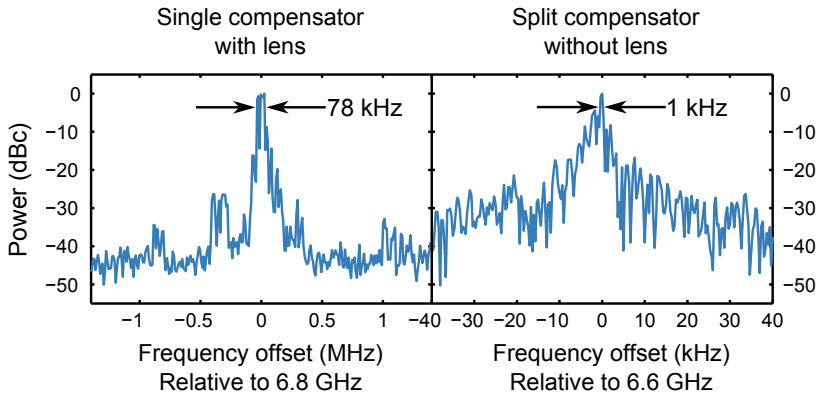


Figure 4-14: RF beatnotes measured from laser bias line, for lensed devices (left) and non-lensed devices (right).

lays for these lines can be estimated using $D_2 \approx \frac{1}{2L}\frac{1}{2\pi}\frac{\Delta\tau_g}{\Delta s}$, or $\Delta\tau_g \approx 4\pi L D_2 \Delta s$. If the pair has spacings $\Delta\nu_1$ and $\Delta\nu_2$, then the difference in these spacings is approximately $|\Delta\nu_1 - \Delta\nu_2| \approx \Delta\tau_g \Delta\nu^2$, or

$$|\Delta\nu_1 - \Delta\nu_2| \approx 4\pi L D_2 \Delta\nu^2 \Delta s \tag{4.20}$$

For $D_2 = 0.1$ ps^2/mm, a repetition rate $\Delta\nu = 6.8$ GHz, and a 5 mm cavity, one finds that two line pairs separated by $\Delta s = 100$ GHz will have differences in their repetition rates of about 29 MHz. Clearly, this is much broader than any of the observed beatnotes, so what is happening? Why are kilohertz fluctuations in the beatnote frequencies still present?

The answer is that the QCLs are in fact acting as frequency combs and that the beatnote indicates this, but they're being broadening by environmental factors, particularly feedback. Because the devices are mounted in a pulsed-tube cryocooler, mechanical vibrations cause the optical feedback to be unstable and impart low-frequency fluctuations on the beatnote. Even when the feedback is blocked by covering the window with an absorptive material, fluctuations remain in the beatnote frequencies, likely because some light reflects off the window itself. Fortunately, by slowly applying a sub-mA modulation to the bias of the laser, most of the phase noise associated with the feedback can be removed and the output can be stabilized. Figure

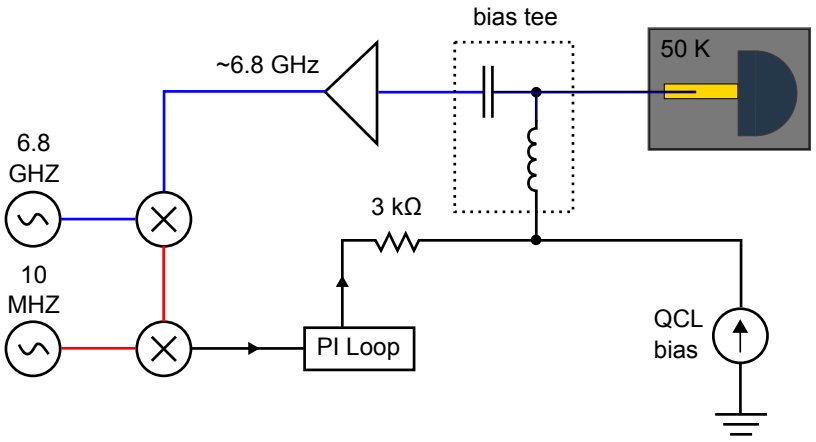


Figure 4-15: Setup used for stabilizing the repetition rate. RF lines are shown in blue (GHz), IF lines are shown in red (MHz), and DC lines are shown in black (kHz).

4-15 shows the setup used for stabilizing the QCL's repetition rate against mechanical vibration. For repetition rate stabilization, the free-running beatnote emanating from the QCL is first observed on a spectrum analyzer, which is typically near 6.8 GHz. An external frequency synthesizer (HP 8673E) is tuned to be 10 MHz away from the free-running signal. The QCL beatnote is then downconverted twice, first to 10 MHz and then to DC, and is used as the error signal for a PI controller. The output of the PI controller is added to the QCL bias with a 3 kΩ resistor, and since the QCL's bias affects the refractive index and therefore repetition rate, this locks the QCL's repetition rate to the frequency of the synthesizer plus 10 MHz. Note that this is not injection locking, and that the added current only has frequency components up to about 100 kHz, and has an amplitude of less than a milliamp (much smaller than the DC bias of about an amp).

Figure 4-16 shows the effect of this stabilization procedure on the beatnote, as measured by downconverting the beatnote with the RF and IF local oscillators and recording the beating on an oscilloscope, and Fourier Transforming to get the power spectrum. (Above 40 MHz the signal is measured with an RF spectrum analyzer instead, and so the signal is discontinuous at 40 MHz.) 99.92% of the beatnote's free-running phase noise is removed, and no major sidebands are visible other than the one at 350 kHz created by the phase-locked loop.

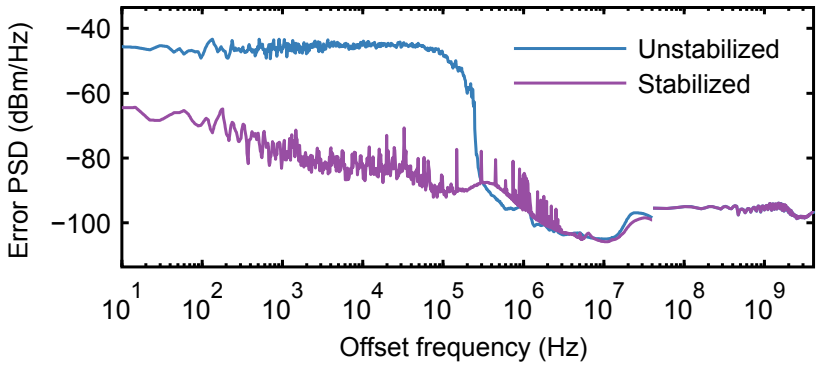


Figure 4-16: Effect of stabilizing beatnote against external perturbations.

Of course, the existence of a broad spectrum does not prove that the QCL is behaving as a frequency comb, merely as an interesting multi-mode laser. Even relying on the beatnote's linewidth and phase noise is questionable, as QCLs are frequently prone to microwave oscillation as a consequence of their negative differential resistance. (Negative differential resistance is even used to construct oscillators, such as in resonant-tunneling diodes.) The following chapter establishes the principles necessary for describing such a laser's coherence, and shows that this device is truly a comb.

Chapter 5

Coherence of frequency combs

Why was the laser fundamentally different from all earlier light sources? There are many ways to answer this question, but perhaps the most succinct is that it is coherent. The word "coherent" is ambiguous and generally depends on the context, but the simplest way to define a coherent system is as something that can be characterized by a single number. For example, a light source is temporally coherent if its time-dependence is completely described by a single frequency. Similarly, a light source is spatially coherent if it emits completely into a single mode, in which case only one phasor quantity is needed to describe its amplitude and phase.

It should therefore be no surprise that if one wants to use a frequency comb for spectroscopic and metrological applications, coherence is probably the most important parameter that needs to be characterized. For frequency combs, there are two types of coherence that can be discussed, corresponding to the two frequencies that describe it. The first is its *mutual coherence*, which describes the relative phase stability of any two lines of the comb. Because phase-locking in laser microcombs is accomplished through four-wave mixing, which in return requires a uniformity in difference frequencies, this is effectively equivalent to the constraint that the lines be evenly spaced. In other words, mutual coherence is related to the uniformity of the comb's repetition rate. The other type of coherence is its *absolute coherence*, which describes the global phase stability of all the lines. Absolute coherence is related to the comb's offset frequency.

Why are these two properties important? Consider the application of dual-comb

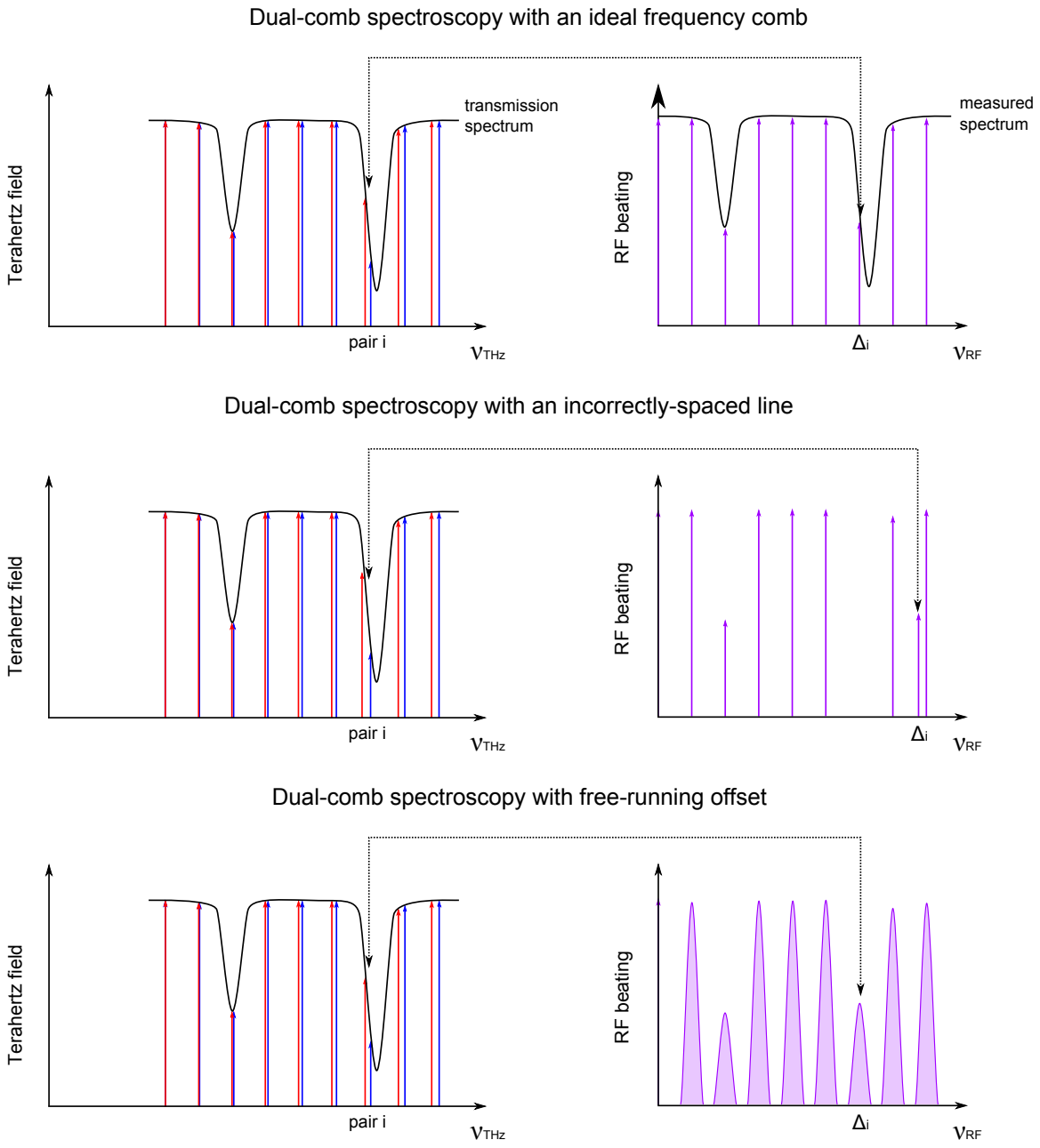


Figure 5-1: Dual comb spectroscopy with various types of incoherence.

spectroscopy, a schematic of which is shown in Figure 5-1. Suppose that the ith pair of comb lines is located at a frequency ν_i. If ideal frequency combs were used to detect the absorption of a gas by shining them through it, then the beating between those lines would be located at the radio frequency Δ_i. By measuring all of the beatnotes between each pair of lines, one could reconstruct the original sample's absorption

profile.

Next, consider the effect of mutual incoherence on the measured absorption. This could manifest as a line which is detuned from the rest of the lines in the comb; here one line is slightly red-shifted compared to the rest. The result of this red-shifting is that the beatnote frequency Δ_i is increased, causing the resulting RF spectrum to be comb-like no longer. This severely complicates an analysis of the resulting data: though one could still conceivably extract the absorption profile from this measurement, doing so would require precise knowledge of both combs' frequencies, a nontrivial task.

Finally, consider the effect of absolute incoherence on the absorption profile. Even though all of the lines are uniformly-spaced, because each line has a finite linewidth and the absolute frequencies of each comb are uncorrelated, the resulting beatnotes will also broaden. If this broadening is too severe, then the beatnotes will overlap and produce uninterpretable data.

5.1 Mutual coherence

5.1.1 Mutual coherence of two lines

To discuss the different types of coherence used in frequency combs, one must mathematically clarify what is meant by each type. It is therefore easier to start with the case of two lines. If an ideal oscillator of frequency ω has an amplitude A and a phase ϕ, then the field associated with that oscillator would be $E(t) = \mathrm{Re}\left(Ee^{i\omega t}\right)$, where the quantity $E \equiv Ae^{i\phi}$ is referred to as its phasor. However, this time-dependence presupposes that the oscillator is completely stable and unchanging with time. In reality, if the oscillator's amplitude and phase were free to drift around, the phasor would acquire an explicit time-dependence. Instead, the electric field should be expressed in the form $E(t) = \mathrm{Re}\left[A(t)e^{i(\omega t+\phi(t))}\right]$. Note that if the amplitude of the oscillator were fixed, then the linewidth of the oscillator will be wholly determined by fluctuations in its phase.

As previously mentioned, mutual coherence is defined in this thesis as the relative phase stability of two lines. In other words, if two lines have negligible amplitude fluctuations and have respective phases of $\phi_1(t)$ and $\phi_2(t)$, then their relative phase difference $\phi_{21} \equiv \phi_2 - \phi_1$ will be a metric for how phase-coherent they are. More specifically, when ϕ_{21} is constant in time, then the oscillators will be considered completely coherent (even if that phase difference is non-zero). This can be further quantified by use of the phase noise power spectral density $S_{\phi_{21}}(\omega)$, which is defined as the power spectrum of the phase fluctuations per unit bandwidth (and has units of rad^2/Hz):

$$S_{\phi_{21}}(\omega) \equiv \left\langle \frac{1}{T} \left| \int_0^T \phi_{21}(t) e^{i\omega t} dt \right|^2 \right\rangle \tag{5.1}$$

In the previous expression, it is assumed that the measurement is repeated many times to get a statistical average, and that the phase noise is recorded over some long interval T, referred to as the measurement time. (One can alternatively conceive of this as an ensemble average over many systems.) Also note that the above quantity is defined over positive ω and is therefore considered a single-sideband measurement. Since the Fourier transform of a real quantity is symmetric, without loss of generality one can alternatively consider only positive frequencies as long as the corresponding phase noise is doubled. Physically, this is a consequence of the fact that both quadratures of a sinusoid are free to carry noise. In addition, it is usually more convenient to deal with linear frequencies rather than angular frequencies, and so the convention used here is that the phase noise plotted is always $S_{DSB}(\nu) \equiv 2S_{\phi_{21}}(2\pi\nu)$.

How can the relative phase be measured? The most straightforward method is simple photodetection. When lines of frequency ω_1 and ω_2 and phase ϕ_1 and ϕ_2 are shined on a sufficiently fast photodetector, they will beat together to form an electronic frequency at $\omega_2 - \omega_1$ with a phase given by $\phi_2 - \phi_1$. Once again, fluctuations in the relative phase will translate into a non-zero difference-frequency linewidth. Next, if an electronic local oscillator is used to further downconvert the difference frequency, then the phase can directly be measured down at sub-MHz frequencies. Figure 5-2 shows the steps of this process.

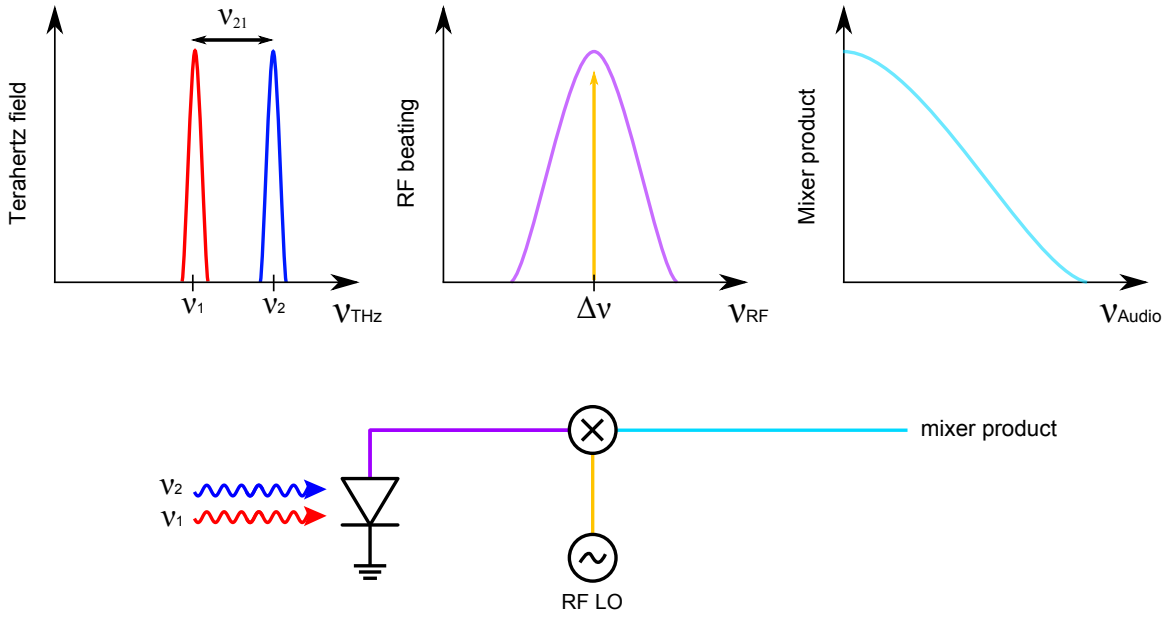


Figure 5-2: Schematic showing the process used to measure phase noise.

More rigorously, suppose that the oscillators have fields $\mathrm{Re}\left[E_1(t)e^{i\omega_1 t}\right]$ and $\mathrm{Re}\left[E_2(t)e^{i\omega_2 t}\right]$. The power measured by the photodetector, $P(t) \sim E^2(t)$, will effectively be low-pass filtered since no photodetector can respond to optical fluctuations in its signal, leaving

$$P(t) = \frac{1}{2}\left(|E_1|^2 + |E_2|^2 + E_1^* E_2 e^{i(\omega_2-\omega_1)t} + E_1 E_2^* e^{i(\omega_1-\omega_2)t}\right). \tag{5.2}$$

The first two terms correspond to usual power detection and are phase-insensitive, while the second two terms are sensitive to the optical fields' phase differences. Next, suppose an electronic local oscillator (LO) with a well-defined amplitude and phase is tuned near the beat frequency $\omega_{21} \equiv \omega_2 - \omega_1$. If the LO's time dependence takes the form $\mathrm{Re}\left[V_0 e^{i\omega_0 t}\right]$, then the resulting mixed product will be $S(t) = P(t) \times \mathrm{Re}\left[V_0 e^{i\omega_0 t}\right]$, once again low-pass filtered:

$$S(t) = \frac{1}{2}\mathrm{Re}\left[V_0 E_1(t) E_2^*(t) e^{i(\omega_0-\omega_{21})t}\right] \tag{5.3}$$

Rewriting this result in terms of amplitudes and phases, one finds that the mixed

product is proportional to

$$S(t) = A_1 A_2 \cos\Big((\omega_0 - \omega_{21})t + (\phi_0 - \phi_{21})\Big). \tag{5.4}$$

In other words, by detecting the beatnote and mixing the result on a local oscillator, we have obtained a slowly-varying electronic sinusoid whose phase is related to the optical phase noise $\phi_{21}(t)$. But how can this be used to get the phase noise directly? One approach is to tune ω_0 to be exactly equal to ω_{21} and to choose $\phi_0 = \pi/2$, in which case the resulting signal would be $M(t) = A_1 A_2 \sin(\phi_{21}) \approx A_1 A_2 \phi_{21}(t)$. If this could be done, the phase noise spectral density could be found by simply recording $S(t)$ and performing a fast Fourier transform (FFT).

Unfortunately, in reality it doesn't matter how closely the local oscillator's frequency is tuned to the beat frequency, since independent oscillators will eventually drift off of each other. Luckily, feedback can be used to remedy this problem. If the phase error ϕ_{21} is not too large and the detuning frequency is sufficiently small, then the total electronic phase will change slowly and negative feedback can be used to ensure that the local oscillator and the beat frequency remain in phase with each other. Suppose that the frequency of the local oscillator can be adjusted by the application of a voltage. At some point, the total electronic phase will evolve to the point where $\Phi(t) \sim \pi/2$, at which point the mixing product will be linear in the phase. At this point, if sufficient negative feedback is applied it will "lock" the frequency of the local oscillator to the beat frequency. If the lock is sufficiently robust, it will ensure that the phase error is massively reduced.

Next, suppose that the local oscillator is now used to measure both quadratures of the beating between lines 1 and 2. For example, this could be accomplished by phase-shifting the local oscillator by 90° and producing a second mixing product. Ignoring the local oscillator's amplitude, and shifting the time origin so that the LO's

phase is 0°, the two mixed signals are proportional to

$$S_I(t) = \text{Re}\left[E_1(t)E_2^*(t)e^{i(\omega_0-\omega_{21})t}\right] \tag{5.5}$$

$$S_Q(t) = \text{Im}\left[E_1(t)E_2^*(t)e^{i(\omega_0-\omega_{21})t}\right] \tag{5.6}$$

where $S_I(t)$ is used to refer to the "in-phase" signal and $S_Q(t)$ is used to refer to the "in-quadrature" signal. Even if one of the quadratures is locked by feedback to zero, the other one will by necessity be non-zero, since the two signals are out of phase with each other (i.e., one or both will acquire a D.C. value). Essentially, one of the quadratures is cosine-like and the other is sine-like. Figure 5-3 shows a schematic of the power spectral densities associated with each quadrature, in both the locked and unlocked states. Note first that locking does not completely eliminate

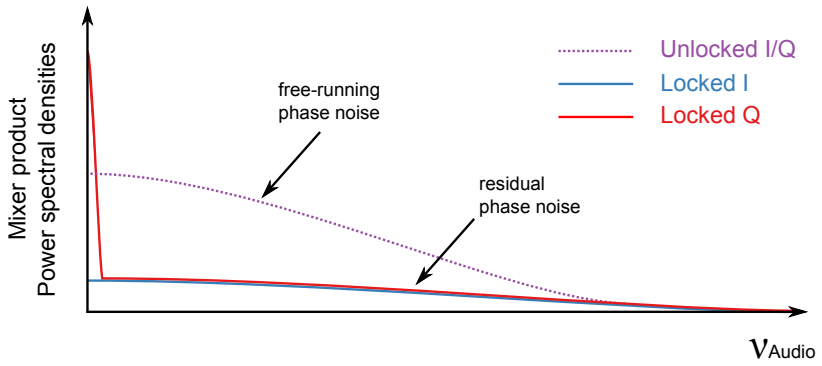


Figure 5-3: The power spectral densities associated with each quadrature of the mixed signals.

the phase noise, as there will always be residual phase noise that can't be elimnated by feedback (although it can be greatly suppressed). Note secondly that the power spectral density of one of the quadratures—in this case the Q channel—contains a delta function at zero frequency, by virtue of its DC term. Moreover, note that the DC term will essentially vanish in the unlocked state, since sufficiently large phase fluctuations will ensure that the long-term average of both quadratures is zero.

A mathematically cleaner way of describing the two quadratures is to combine them into a single signal that is complex, S_+. While this can't be done physically, it

can be done in a computer provided both quadratures are simultaneously recorded. In other words, the following definitions are made:

$$\begin{aligned} S_{\pm}(t) &\equiv S_I(t) \mp iS_Q(t) \\ S_+(t) &= E_1^*(t)E_2(t)e^{i(\omega_{21}-\omega_0)t} \\ S_-(t) &= E_1(t)E_2^*(t)e^{-i(\omega_{21}-\omega_0)t} = S_+^*(t) \end{aligned} \tag{5.7}$$

Note that the quantity S_- has also been defined, representing S_+'s conjugate. When the beatnote between the lines has been locked to ω_0 using feedback, S_+ ideally becomes constant in time. Moreover, the phase of S_+ is then nothing more than the relative phase of the two lines. As a result, fluctuations in the relative phase manifest as fluctuations in S_+.

Next, consider the average value of the complex signal, $\langle S_+(t)\rangle$. Here, angle brackets refer to a time average over long-term (millisecond or longer) time scales. To see how this quantity is affected by phase fluctuations, it is useful to plot the complex beating on a phasor diagram, as shown in Figure 5-4. On these diagrams,

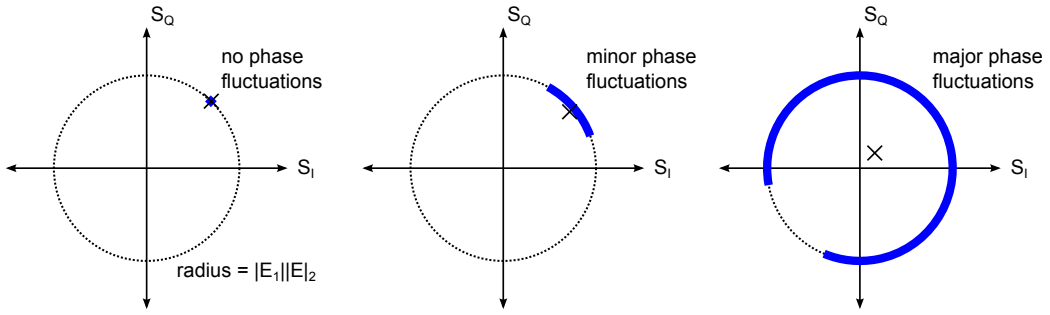


Figure 5-4: Phasor diagram for varying levels of phase fluctuation. The average value of S_+ is marked with an x.

$S_+(t)$ has been drawn with a blue line, its mean value has been marked with an $\times$, and the circle representing $|S_+| = |E_1||E_2|$ has been drawn for reference. Ignoring amplitude fluctuations, if the beat signal between the lasers is completely phase-stable, as shown in the left-most case, then the mean value of $S_+(t)$ lies on the circle. If the phase fluctuations are only minor, as shown in the middle case, then the mean

value will lie on the circle's interior. However, if the phase fluctuations are major, as shown in the right-most case, the mean value will be close to the origin. As the magnitude of the fluctuations approach 2π, the magnitude of the mean essentially vanishes.

It is with this picture in mind that mutual coherence is defined. Normalizing $\langle S_+(t)\rangle$ to the amplitudes of lines 1 and 2, the mutual coherence is defined to be

$$g_+(\omega_{21}) \equiv \frac{\left|\left\langle E_1^*(t)E_2(t)e^{i(\omega_{21}-\omega_0)t}\right\rangle\right|}{\sqrt{\langle|E_1(t)|^2\rangle\,\langle|E_2(t)|^2\rangle}} \tag{5.8}$$

This definition of coherence is convenient because it can be directly measured: mixing can be used to measure the numerator, and power measurements can be used to measure the denominator. This definition has also been used by many groups developing microcombs, since notions of timing jitter that work well for mode-locked lasers are not as relevant for microcombs.[98] Note that this coherence metric is dependent on the choice of offset frequency ω_0: when $\omega_{21} \neq \omega_0$, this coherence will vanish and $g_{21} = 0$. When there are no long-term phase fluctuations between the two lines and the frequencies are spaced by ω_0, they will be completely coherent and $g_{21} = 1$.

5.1.2 Mutual coherence of comb lines

The generalization of this definition to the multiple lines of a frequency comb is straightforward. One can say that the coherence with respect to the offset ω_0 between lines i and j is

$$g_+(\omega_{ji}) \equiv \frac{\left|\left\langle E_i^*(t)E_j(t)e^{i(\omega_{ji}-\omega_0)t}\right\rangle\right|}{\sqrt{\langle|E_i(t)|^2\rangle\,\langle|E_j(t)|^2\rangle}} \tag{5.9}$$

Though writing this definition down is easy, actually measuring g is challenging. While it is possible in principle to perform a mixing experiment similar to the one previously described, one would have to measure the coherence for each pair of lines. In practice, not only would this be extremely difficult (requiring N^2 measurements for N lines), but also effectively impossible since the largest difference frequency involved in the spectra previously shown is about 800 GHz, which would require a terahertz

detector with a bandwidth of 800 GHz. Even if such a detector and IF circuitry could be found, each line would have to be isolated, requiring optical bandpass filters with a bandwidth much less than the repetition rate, 6.8 GHz.

Of course, proving the existence of a frequency comb whose lines are evenly spaced does not actually require that every value of $g_+(\omega_{ji})$ be measured. In fact, one only needs to show that every line is separated from its neighbors by the same frequency, the repetition rate $\Delta\omega$. For an ideal frequency comb, the coherence between lines n and $n+1$ with respect to the repetition rate, $\Delta\omega$, should be equal to 1 for all n. This motivates an alternative definition for the coherence of a comb:

$$g_+(\omega) \equiv \frac{|\langle E^*(\omega)E(\omega+\Delta\omega)\rangle|}{\sqrt{\langle|E(\omega)|^2\rangle\,\langle|E(\omega+\Delta\omega)|^2\rangle}}, \tag{5.10}$$

This definition plays loose with the admixture of time and frequency, but it can be made rigorous by using short-time Fourier Transforms and taking an average over macroscopic timescales. In this definition, the frequency detuning factor $e^{i(\omega_0-\omega_{ij})t}$ has effectively been wrapped into the frequency-shifting factor $\omega+\Delta\omega$. If a line at ω has no line next to it at exactly $\omega+\Delta\omega$, the coherence will vanish.

5.2 Beatnote measurements of mutual coherence

Unfortunately, reducing the number of beatnotes that need to be measured does not really solve the problem of how $g_+(\omega)$ can be measured, because spectral filtering would still be necessary to obtain its frequency dependence. One *could* shine all of the modes onto a single detector without prefiltering, and this would give some information about the individual beatnotes since they would all appear on the RF spectrum analyzer, but such a measurement is incoherent and gives limited information about g. The narrowness of the beatnote is only a necessary condition to show that a laser is acting as a comb, but it is not a sufficient one. Still, such measurements provide a useful check and are performed here.

As mentioned earlier, an interesting side effect of the broadband terahertz gener-

ation in properly-compensated frequency comb devices is that they generate a strong RF beatnote at the round-trip frequency of the laser when DC-biased. Moreover, this beatnote is narrowband (with a full-width half maximum of about 100 kHz) when compared with its center frequency (6.8 GHz), suggesting that it is extremely coherent. But what does it represent, and is it physical? One possibility is that the QCL is acting as a mixer of sorts for the light circulating inside the cavity. This intracavity light is spaced by the mode spacing of the cavity and therefore generates beating much like a mixer does.

There are several ways for this to happen, which are mathematically indistinguishable and yet physically very different. The first way is essentially direct detection. It is known that the current drawn by a QCL is a relatively strong function of the circulating power, since the upper state lifetime of a laser is sharply reduced by stimulated emission. This can be seen by comparing the lasing IV of a laser with the non-lasing IV of the same laser, an example of which is shown in Figure 5-5. (Lasing

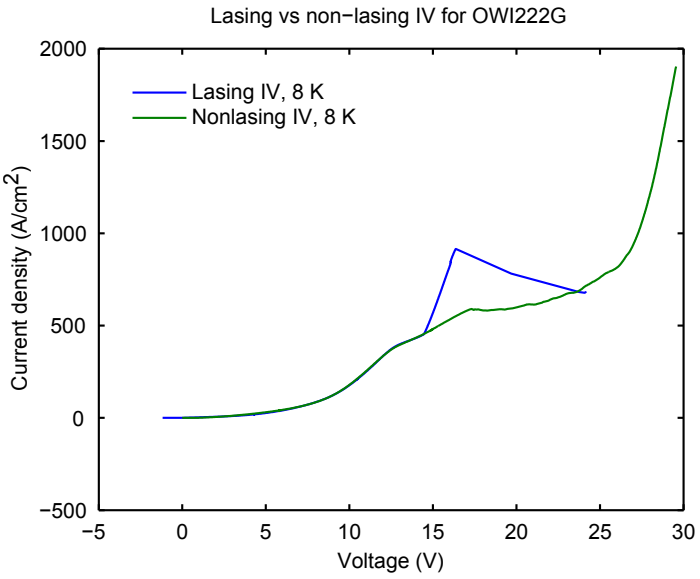


Figure 5-5: Lasing current-voltage vs nonlasing current-voltage for a laser comprised of the OWI222G gain medium [28].

in this structure was inhibited by the application of Stycase 2850FT, an epoxy which is very lossy in the terahertz.) Notice that the presence of a strong optical field inside the cavity increases the current density through the laser by almost a factor of 2.

Fundamentally speaking, this means that the QCL *is* essentially a detector, albeit one with a relatively low responsivity. How fast is it as a detector (i.e., how quickly does it respond to changes in its input)? If treated as a traveling-wave detector, it should be as fast as its gain recovery time, on the order of tens of picoseconds for terahertz QCLs. [99] Since the metal-metal waveguide is essentially a microstrip line that supports microwave frequencies in addition to terhertz ones, it would be no surprise if the gain medium were able to transmit local fluctuations in intensity into fluctuations in the local bias of the waveguide that would be measurable with an external microwave spectrum analyzer.

Another possibility that would explain the generation of difference frequencies inside the waveguide is the possibility of Schottky mixing on the contacts of the structure. Because the highly-doped layer needed to form an ohmic contact between the gain medium and its metallic contact is moderately lossy in the terahertz, it is often the case that the contact layer is removed and a Schottky contact remains. As Schottky diodes can be used as terahertz detectors [53] by virtue of their high-speed electronic nonlinearity, Schottky mixing can be used to detect intracavity difference frequency mixing. For example, such behavior has been observed in structures that were deliberately engineered to act as Schottky mixers by Wanke et al. [100]. The last possibility is simple difference frequency generation by means of optical rectification (i.e., by $\chi^{(2)}$ second-order nonlinearity). As intersubband structures can have extremely large nonlinearities [101], frequency mixing can act to form the difference frequency. But whatever the mechanism, be it direct detection, photonic, or electronic, all produce signals at the difference frequencies inside the cavity, which can be measured outside the laser using a bias tee and an RF spectrum analyzer.

5.2.1 Schottky mixer comparison

In order to check how representative this measurement is of the "real" beatnote that would be observable on an external detector, we compared the QCL beatnote to the beatnote measured on a fast THz detector—initially, a horn-coupled Schottky mixer from Virginia Diodes (VDI). The VDI Schottky mixer can respond to terahertz signals

more or less instantaneously in a compact room-temperature form factor (meaning that their effective IF bandwidth is whatever frequency they're operated at), but do so at the expense of a relatively low sensitivity. Figure 5-6 shows a comparison

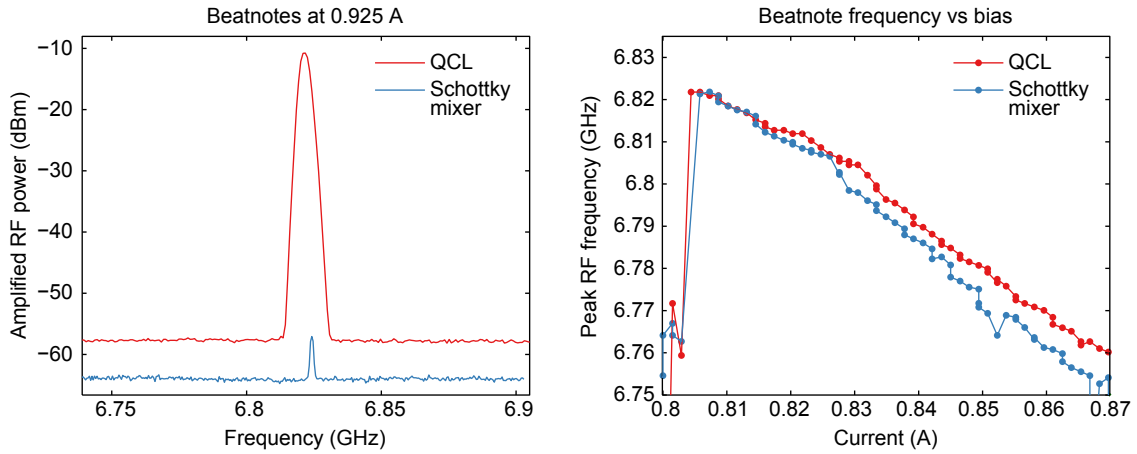


Figure 5-6: Left panel: Beatnote obtained directly from a lens-coupled QCL and from a Schottky mixer at 45 K and a bias of .925 A. Right panel: Bias dependence of the beatnote, as measured on a Schottky mixer and on a QCL.

between the RF beatnotes from a lens-coupled QCL comb (FL183S-FC-6-m6) and from a Schottky mixer used to detect the comb. (Note that in the left panel the measurements were not taken with the same resolution bandwidth and the linewidth of each beatnote is determined by the instrument.) The signal-to-noise ratio of the QCL-generated beatnote is extremely high—over 50 dB—and is in fact determined by the dynamic range of the spectrum analyzer. However, the signal-to-noise ratio of the Schottky-detected beatnote is only about 10 dB (even though the laser is a powerful 5 mW laser), owing to the diode's relatively low sensitivity and the fact that the QCL measurement is a measure of the intracavity power and should be much higher than emanated power (which suffers coupling losses upon exiting the cavity, passing through free space, and going into the Schottky mixer's horn). Also note that the two beatnotes have a similar bias dependence and are probably actually identical, differing only because only one spectrum analyzer channel was available and the two signals had to be switched manually, causing the environment to drift slightly between the two measurements.

5.2.2 Hot electron bolometer

Another detector that can be used to measure terahertz beatnotes is a hot electron bolometer. Like all bolometers, it detects light by detecting the temperature change in a thermal reservoir induced by incident radiation. Bolometers are particularly well-suited for terahertz detection since they have no intrinsic frequency cutoff that would limit operation at terahertz frequencies (e.g., nothing equivalent to the bandgap of a photodiode). Though bolometers can be quite sensitive, they are also usually quite slow since they are based on thermal effects. Very few types of bolometers can achieve gigahertz speeds. Hot electron bolometers (HEBs), on the other hand, function by coupling light into a superconducting bridge using an antenna. If the bridge's electrons heat to above the critical temperature of the material, superconductivity is lost and the bolometer becomes a resistor. This change in resistivity is very dramatic and allow for very sensitive detectors to be made (NEP$\sim 10^{-14}$ W/$\sqrt{\text{Hz}}$). In addition, because in HEBs it is only the bridge's electrons that are heating up, the system has a very low heat capacity. As the electron-phonon interaction efficiently transfers energy into the lattice, the electronic system has a very short thermal time constant (sub-ns). This endows the HEB with an excellent combination of both sensitivity and speed at a wavelength where both are usually lacking. Their main tradeoff is that they usually require cooling to liquid helium temperatures.

For these measurements, a superconducting NbN HEB mixer obtained from collaborators at TU Delft and at SRON in the Netherlands was used. This device in particular is known for its extreme sensitivity [55] and response beyond 7 GHz [102], and has proven to be a very useful tool for performing experiments with THz QCLs. Figure 5-7 shows a schematic of this HEB, as well as a picture of how it was mounted in a helium dewar. The cryostat was constructed to have high-density polyethylene (HDPE) windows, and a metal mesh low pass filter was used to block unwanted far-infrared radiation from saturating the instrument. The HEB mixer block was mounted close to the window, and an RF bias tee was placed inside the cryostat along with a SiGe cryogenic low-noise amplifier (LNA) that provided 27 dB of gain at

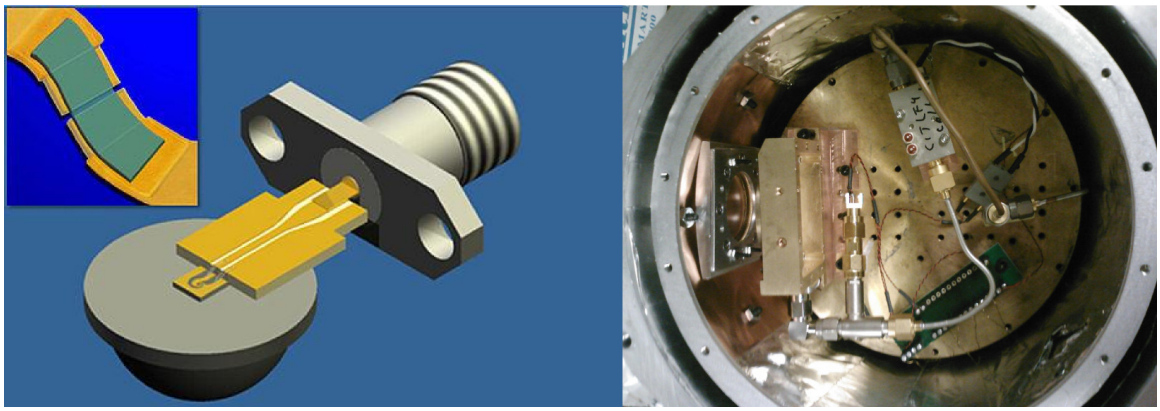


Figure 5-7: Left panel: HEB schematic. Right panel: HEB configuration.

4 K. Stainless steel coaxial cable was used to collect the amplified RF from the LNA, and a custom-built bias box was used to generate the mV-level bias for the HEB as well as measure its DC current. Figure 5-8 shows the DC IVs for the HEB under various pump states. When the HEB is unpumped, the conductivity of the HEB is effectively infinite (for small biases). As more teraherz power is put into the element, the IV curve becomes more and more linear until it effectively becomes resistive. Although this makes the HEB very sensitive, it also makes it prone to saturation. Frequently, a screen had to be used for the HEB to produce meaningful data.

As previously discussed, a straightforward way to test whether a laser is operating as a frequency comb is to shine its light onto a fast detector and to examine the resulting optical beating between all the laser modes. If only one narrow beatnote was observed, that would indicate that all of the detected modes were evenly spaced. Conversely, a non-comb would have multiple beatnotes; e.g, two frequencies separated by 100 GHz would have mode spacings differing by 29 MHz as calculated before.

Figure 5-9 shows various RF beatnotes measured using the HEB, offset from the repetition rate of 6.8 GHz. All beatnotes were measured by attenuating the light from the QCL by 16 dB to prevent saturation (a 40$\times$ reduction), amplifying the RF signal with the 27 dB gain cryogenic LNA, and amplifying the RF further with a 31 dB room-temperature LNA. In all cases, the SNR of the beating seen is far better than the SNR

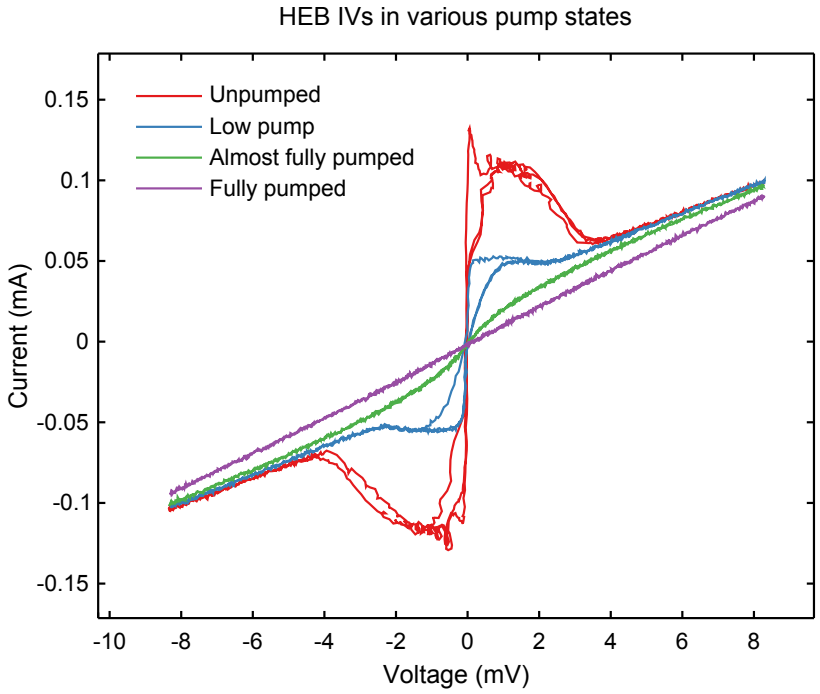


Figure 5-8: IV curves of HEB under various pump conditions.

of the beating seen on the Schottky mixer, and is in excess of 50 dB. The leftmost panel shows the kind of HEB beatnote achievable from a QCL with a lens whose QCL beatnote is unstabilized: the short-term FWHM is comparable to the FWHM of the unstabilized QCL beatnote. The second panel shows the effect of stabilizing the QCL beatnote on the HEB beatnote. Even though feedback is being used to lock the *QCL* beatnote, the *HEB* beatnote also shows a collapse in its linewidth. In this case, it collapses to the spectrum analyzer's minimum resolution, showing that the RF beating coming out of the QCL—generated by whatever mechanism—is a fair representation of the beatnote on the QCL. As a sanity check for this, Figure 5-10 plots the beatnotes from the QCL and from the HEB. Once again, they agree very well, with the differences only being accounted for drift in the meausrement. Lastly, in the last panel of Figure 5-9 the HEB beatnote is downconverted and measured with an FFT spectrum analyzer to get around the spectrum analyzer's minimum resolution. When not limited by the instrument, the beating is only broadened by Hz levels. Though there are sidebands at 60 Hz, these are due to power line fluctuations.

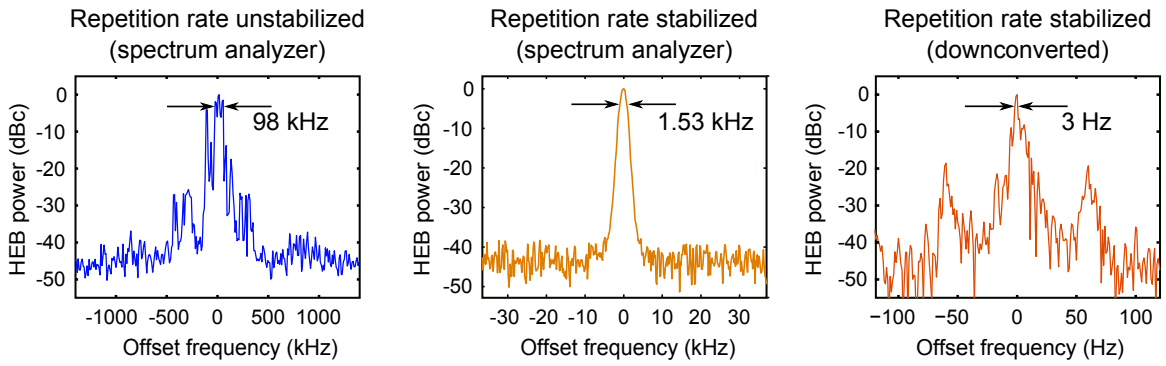


Figure 5-9: RF beatnotes measured using HEB.

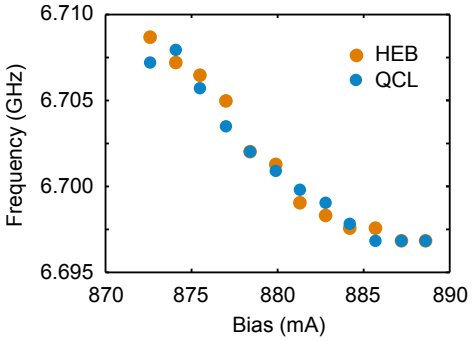


Figure 5-10: Beatnotes measured from HEB and from the QCL versus bias.

This implies that the modes contained within the beatnote are coherent over 10^9 round-trips through the laser cavity [103].

5.2.3 Non-comb biases

It is worth noting that not all biases of the laser form a narrow beatnote, and the output cannot be considered a comb at these biases. Figure 5-11 shows some of these, along with the beatnote standard deviation versus bias in the lasing regime. The bias dependence of the beatnote linewidth shows several regions where the linewidth drops to the resolution bandwidth of the spectrum analyzer (potentially frequency combs), as well as several regions where the linewidth is quite broad (not frequency combs). The leftmost beatnote, indicated by the blue dot on the lower plot, is an example of a "clean" beatnote, exhibiting no sidebands and only possessing a single stable line. The

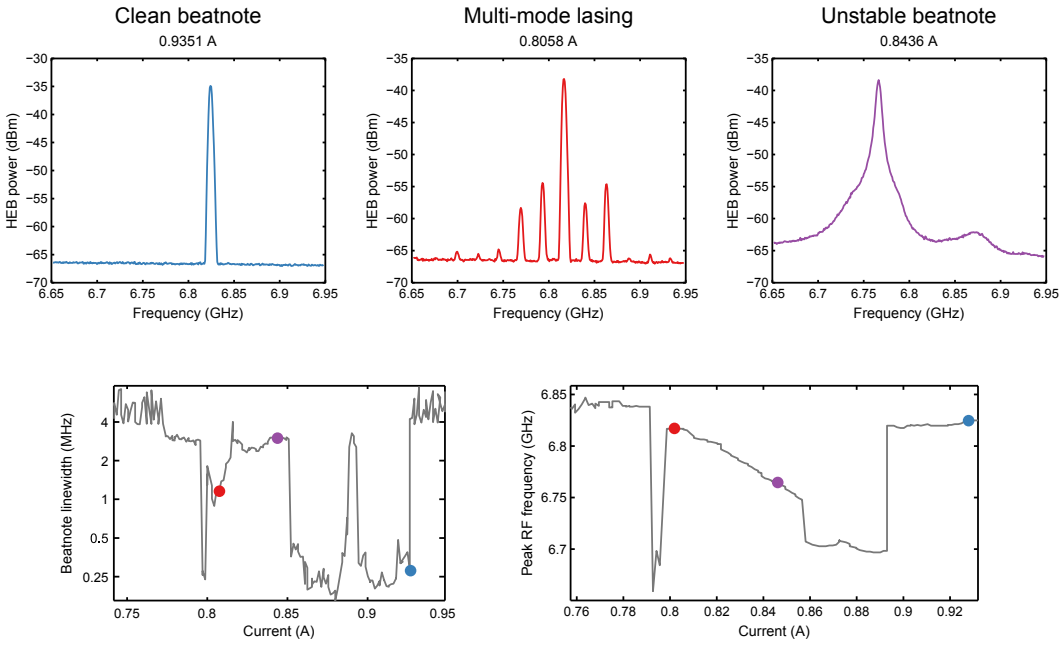


Figure 5-11: Top: Beatnotes measured using HEB at biases which are not combs. Bottom: Beatnote standard deviation and center frequency versus bias.

middle beatnote, denoted by the red dot, is narrow, but also has discrete sidebands which are approximately evenly spaced. The spacing of these lines was found to be bias-dependent, and they are attributed to the laser entering a multi-mode regime in which several lines are lasing, but the four-wave mixing process is not sufficient to phase-lock them and no comb is formed. The last beatnote, denoted by the purple dot, is possesses two peaks and is very broad (10s of MHz). This beatnote is chaotic and is attributed to the competitive nature of the four-wave mixing process. Since the gain spectrum possess two lobes, in the absence of perfect dispersion compensation they will each prefer to form combs at different repetition rates. In stable regions one repetition rate is dominant and wins out, but in unstable regions they continuously fight each other, leading to a broad beating. Note that essentially all of the instability arises in regions where the beatnote frequency is changing quickly with bias.

5.3 Shifted Wave Interference FTS (SWIFTS)

5.3.1 Basic principles

A narrow beatnote is a necessary but not a sufficient condition to prove that a source is a coherent comb. For example, even if a beatnote was narrow it might be the case that a strong pair of lines simply dominate it. It could be the case that some of the lines aren't being detected for some reason, or that their beatnote is so broad that they are buried in the noise floor of the spectrum analyzer. A hypothetical way around this problem would be to use a narrow spectral filter (like a diffraction grating plus a slit) to measure the beatnote as a function of wavelength. Such a scheme is shown in Figure 5-12.

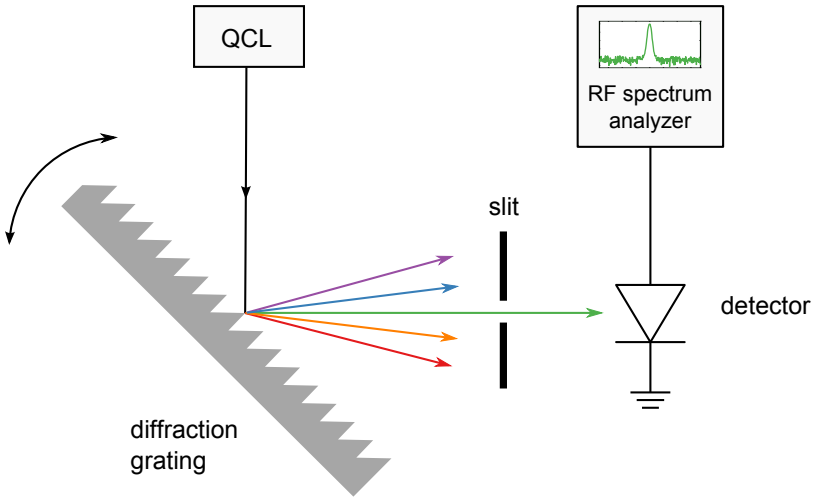


Figure 5-12: Hypothetical setup utilizing spectral filtering to measure the coherence.

Nevertheless, this type of setup is problematic for several reasons. For one, the QCL was found to be very sensitive to optical feedback, and changing the orientation of a nearby grating could change the feedback as well as the laser's spectrum. Especially problematic is that changes in feedback often accompany changes in the repetition rate, which would completely defeat the purpose of this type of measurement (since we would like to see whether the repetition rate is a function of frequency). Another problem is that measuring the coherence $g_+(\omega)$ of the comb requires that

all of the lines be individually resolved by the grating. This would require a narrow slit with a limited etendue and would severely impact the signal-to-noise ratio of the measurement. In fact, this is the very reason Fourier Transform Spectroscopy (FTS) is typically used at long wavelengths instead of grating spectrometers.

Up until now the issue of the terahertz spectrometer has been largely ignored. The THz QCL spectra shown in the previous chapters were all measured using FTS, which is essentially the standard way of measuring the spectrum of non-TDS THz sources. FT spectrometers operate on the classical principle that Michelson interfer-

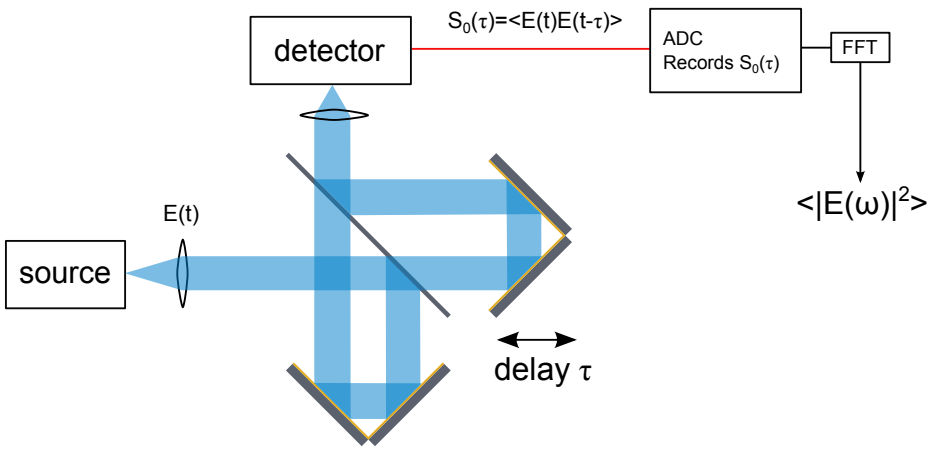


Figure 5-13: Convention FTS schematic. A computer records the autocorrelation of the field.

ometers with variable delays essentially map out the autocorrelation of the electric field, $S_0(\tau) = \langle E(t)E(t-\tau)\rangle$. This is illustrated in Figure 5-13. By recording this signal, known as an interferogram, and Fourier Transforming the result, one obtains the optical spectrum, $S_0(\omega) = |E(\omega)|^2$. Because FT spectrometers measure all of the power in a beam at once, they can provide better signal-to-noise ratios than grating spectrometers at high resolutions. This is the so-called multiplex advantage. It is important to stress that by themselves, conventional spectroscopic techniques like FTS cannot be used to determine the presence of a comb, because their resolution is limited by the physical travel of the mechanical delay element and can only measure spectra to a resolution of about a GHz.[1] In short, while beatnote measurements have

[1]Technically, one can zero-pad the interferogram and achieve better frequency *accuracy* than the

excellent resolution when it comes to measuring difference frequencies, they lack the ability to tell what those frequencies *are*. FTS can tell you what frequencies you have, but cannot tell you with great accuracy what their differences are. This raises an interesting question: is it possible to construct a measurement that combines the high-resolution and throughput of FTS with the precision of beatnote measurements in a single experiment?

The answer is yes, and the result is something the author calls Shifted Wave Interference FTS, or SWIFT spectroscopy [104]. The repetition rate of the QCL is stabilized to a synthesizer as previously described, and the light is passed through a Michelson interferometer as in FTS. Instead of only recording the intensity of the light through the interferometer or of the beatnote, the average value of the in-phase and in-quadrature components of the intensity at the repetition rate are recorded. Figure 5-14 illustrates this. Physically, SWIFTS constitutes a measurement of

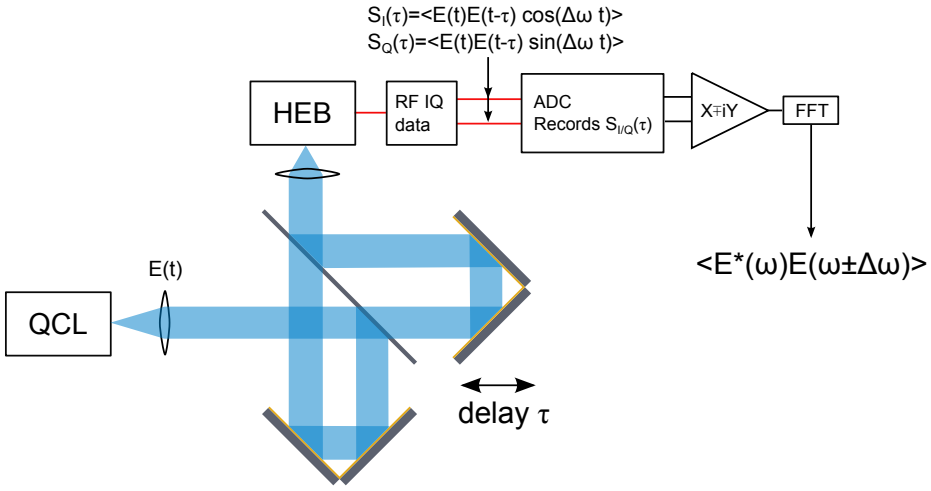


Figure 5-14: SWIFTS schematic. A computer records the quadratures of the auto-correlation.

$$S_I(\tau) = \langle E(t)E(t-\tau)\cos(\Delta\omega t)\rangle \tag{5.11}$$

$$S_Q(\tau) = \langle E(t)E(t-\tau)\sin(\Delta\omega t)\rangle \tag{5.12}$$

resolution. However, this requires that the position of the delay stage be very well-calibrated, and since most commercial FTSs use an unstabilized HeNe laser as their position reference, the resulting precision is not typically much better than the resolution.

As usual, it is more convenient to deal with the complex version of these quantities,

$$S_{\pm}(\tau) = S_I(\tau) \mp iS_Q(\tau) \tag{5.13}$$

The key advantage of SWIFTS is that the Fourier Transform of the analytic functions are almost exactly the quantity previously defined as the coherence:

$$S_{\pm}(\omega) = \langle E^*(\omega)E(\omega \pm \Delta\omega)\rangle\,. \tag{5.14}$$

Here, angle brackets are used to represent both an average over laboratory timescales (seconds) and also a convolution with the instrument's apodization function. This can be easily shown using elementary Fourier theory; for more details, see Appendix C. With this notation, conventional FTS measures $S_0(\omega) = \langle E^*(\omega)E(\omega)\rangle$.

There are some key differences between conventional FTS and SWIFTS. For incoherent light sources, the SWIFT spectrum $|\langle E^*(\omega)E(\omega+\Delta\omega)\rangle|$ vanishes altogether, since the long-term phase incoherence between $E(\omega)$ and $E(\omega+\Delta\omega)$ causes their product to integrate to zero over lab time scales of seconds. Likewise, for a source consisting of two laser lines separated by ω_{21}, the SWIFT spectrum is zero if $|\omega_{21}-\Delta\omega|$ is larger than the integration bandwidth (Hz). One *could* use a normal FT spectrum to define a "spectrum product" as $S_{sp}(\omega) \equiv \sqrt{\langle|E(\omega)|^2\rangle}\sqrt{\langle|E(\omega+\Delta\omega)|^2\rangle}$, and while this product is superficially similar to $|\langle E^*(\omega)E(\omega+\Delta\omega)\rangle|$, it is actually very different. In fact, it will be non-zero as long as $|\omega_{21}-\Delta\omega|$ is within the spectrometer's resolution (GHz). We can therefore compare $|S_+(\omega)|$ and $S_{sp}(\omega)$ to reveal how comb-like a given laser spectrum is. Indeed, the definition of mutual coherence offered in equation 5.10 is simply

$$g_+(\omega) \equiv \frac{|\langle E^*(\omega)E(\omega+\Delta\omega)\rangle|}{\sqrt{\langle|E(\omega)|^2\rangle\,\langle|E(\omega+\Delta\omega)|^2\rangle}} = \frac{|S_+(\omega)|}{S_{sp}(\omega)}, \tag{5.15}$$

SWIFTS therefore provides a direct measurement of the coherence of a source, to Hz levels. It is worth stressing that SWIFTS does not actually improve the spectral *resolution* of the measurement beyond what FTS can normally achieve, since the final

signal is still apodized by the finite delay of the stage. This is illustrated in Figure 5-15, in which two pairs of lines are shown, a red pair and a blue pair. Suppose that

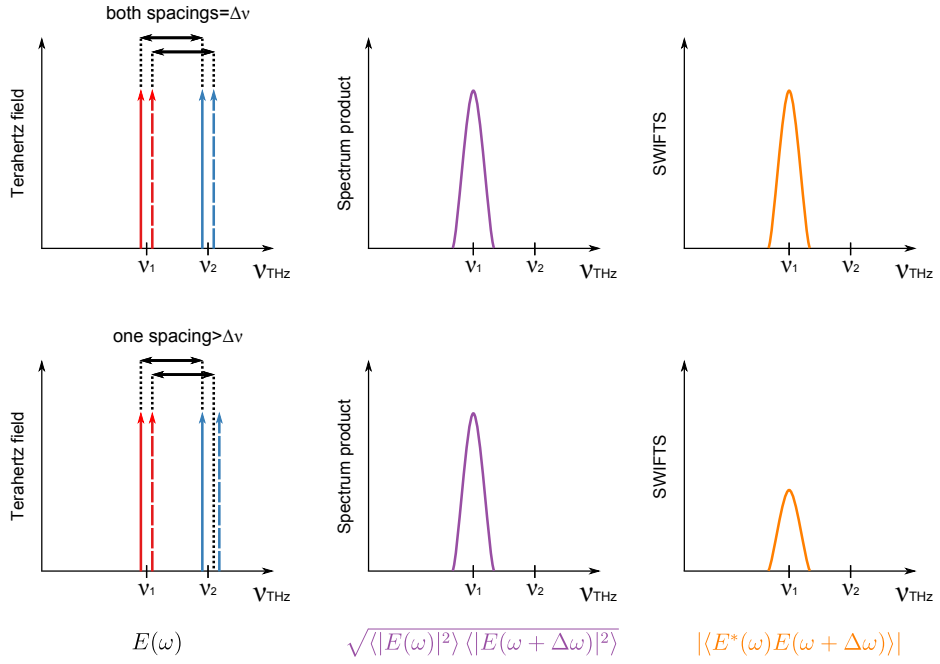


Figure 5-15: Effect of finite apodization on SWIFTS measurement.

the spacing of each pair is very small, say 1 MHz. Because this is far closer than the instrument can resolve, as far as both conventional FTS and SWIFTS are concerned each pair looks like one line. First, consider the case in which the red and blue pairs are exactly $\Delta\omega$ apart, as shown on the top. In this case, the spectrum product and the SWIFTS magnitude produce the same result, and $g = 1$. However, note that each spectrum still look likes a single line. Next, consider the case in which one of the blue lines is slightly detuned from $\Delta\omega$ from its corresponding red line. The spectrum product would be the same as before, but the SWIFTS magnitude will be reduced since the dotted lines no longer contribute to SWIFTS. In this case, $g < 1$.

5.3.2 Key SWIFT results

Figure 5-16 shows the actual system used for most of this analysis. Light from the lens-coupled THz QCL comb was collimated by an f/2 off-axis parabolic mirror (OAP),

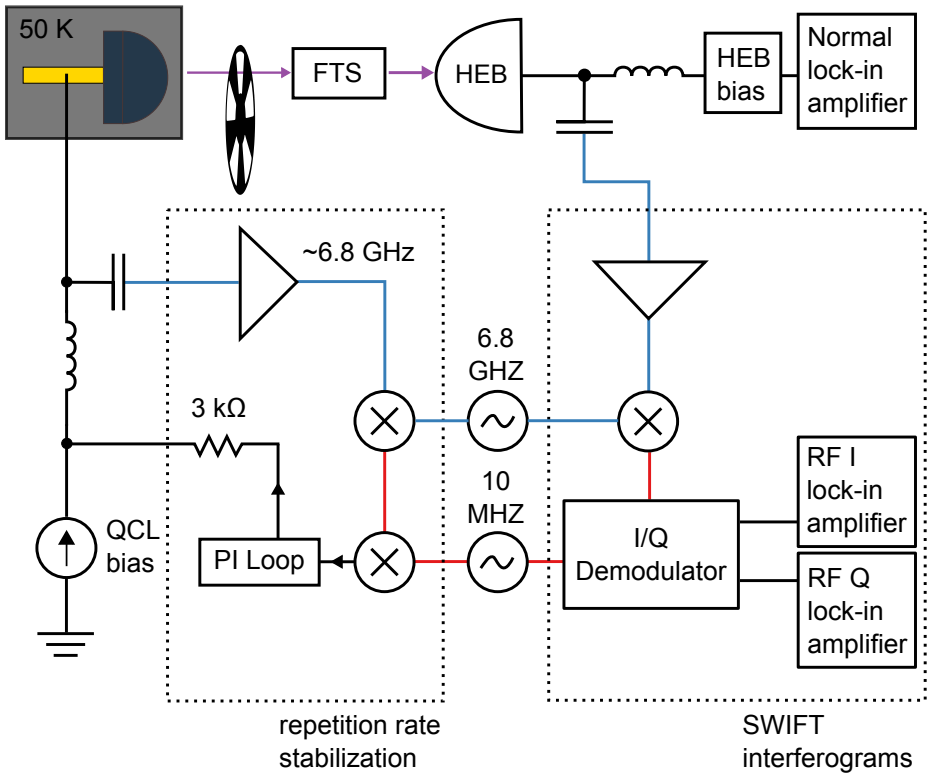


Figure 5-16: Actual implementation of SWIFTS used for most of this analysis.

attenuated by a factor of 40 using a screen to prevent HEB saturation, and passed through an optical chopper. It was then passed through a custom-built nitrogen-purged FTS containing roof mirrors, thereby minimizing retroreflection into the QCL. Lastly, light was focused onto the HEB using an f/3 OAP. Electronically, the HEB was biased to 1 mV using a custom bias box, and the current monitor was passed into a lock-in amplifier to measure the normal interferogram. The RF signal from the HEB was amplified, downconverted to 10 MHz,[2] demodulated with a 10 MHz I/Q demodulator, and passed into lock-in amplifiers. Note that the I/Q demodulation can be performed at 10 MHz rather than near 6.8 GHz since the action of two successive

[2]Higher-order harmonics of the repetition rate can also be used, in which case next-nearest neighbor coherences are being considered.

mixers can be decomposed into the sum of two sinusoids:

$$\cos(\omega_1 t + \phi_1)\cos(\omega_2 t + \phi_2)$$
$$\sim \cos\left[(\omega_1 - \omega_2)t + (\phi_1 - \phi_2)\right] + \cos\left[(\omega_1 + \omega_2)t + (\phi_1 + \phi_2)\right].$$

As the beatnote is locked to just one of $\omega_1 + \omega_2$ or $\omega_1 - \omega_2$, the phase shift of Q relative to I will still be 90°. All three lock-ins used the same time constant, amplitude, and phase settings, and as the Michelson interferometers stage was scanned, their signals were simultaneously recorded. The FTS travel range was 10.75 cm, resulting in a resolution of 1.5 GHz, and the stage speeds varied but were on the order of 0.1 mm/s.

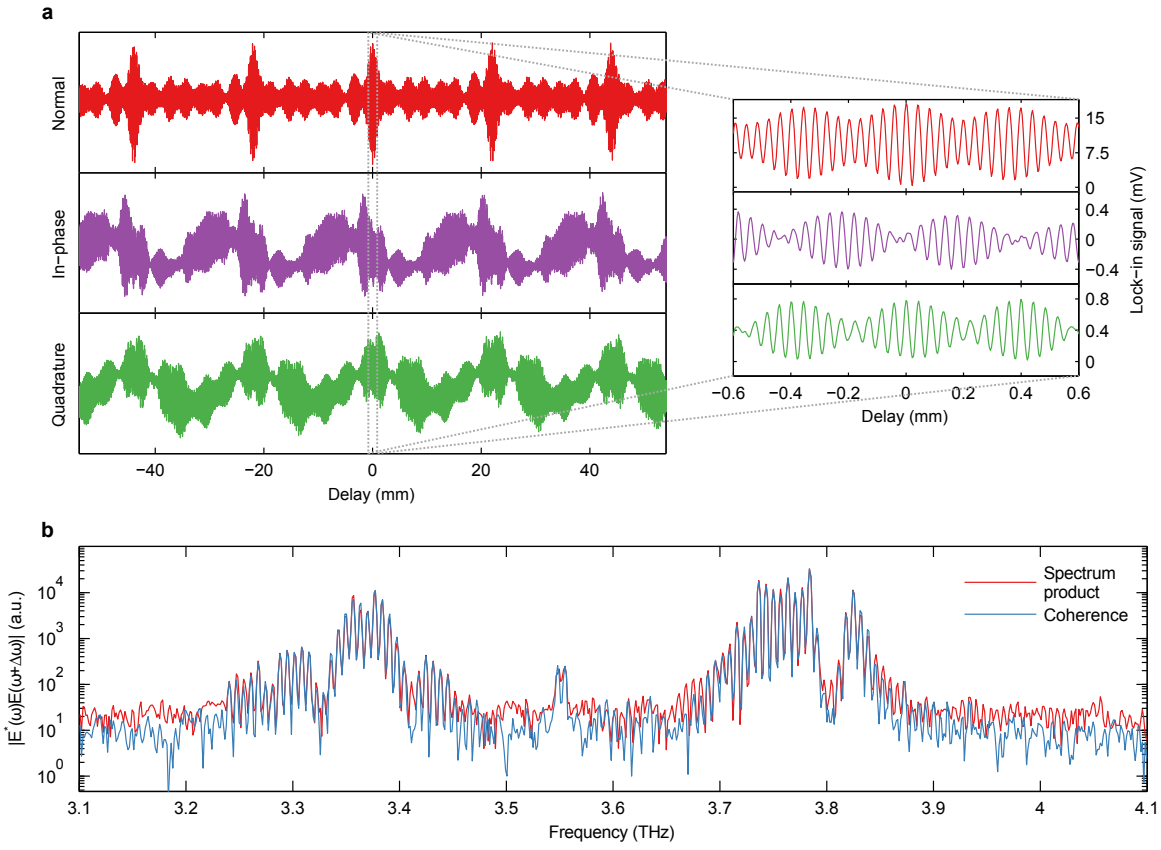


Figure 5-17: Key SWIFTS results for a QCL comb biased to 0.9 A at 50 K. (a) Normal, in-phase, and quadrature interferograms. (b) Frequency domain spectrum product and coherence.

Figure 5-17(a) shows the SWIFT interferograms measured from a device biased

to 0.9 A, along with a normal interferogram obtained at the same bias. Note that while all three interferograms are nearly periodic with a spacing determined by the repetition rate of the laser, the SWIFT interferograms have non-zero phase and as a result are asymmetric about the zero path difference. Next, Figure 5-17(a)(b) shows the SWIFTS coherence magnitude plotted alongside the spectrum product, calculated using FTS. Despite the fact that they are fundamentally different measurements—one coming from the bottom two interferograms and the other coming from the top one—their excellent agreement shows that essentially all of the laser's lines are separated by the same repetition rate and that this can finally be called a frequency comb. The coherence measurement shows approximately 70 comb lines above the noise floor, which span a total range of almost 500 GHz.

It is also worth mentioning that the SWIFTS measurement provides a way to check its own consistency. Ideally, the quantities $S_+(\omega) = \langle E^*(\omega)E(\omega+\Delta\omega)\rangle$ and $S_-(\omega) = \langle E^*(\omega)E(\omega-\Delta\omega)\rangle$ are actually trivially related to each other, since $S_-(\omega+\Delta\omega) = S_+^*(\omega)$. In reality, this is will not exactly be the case since they effectively represent independent measurements and are corrupted by noise. Even so, it should be the case that $S_+(\omega)$ and $S_-(\omega)$ have magnitudes that closely resemble each other if one of them is shifted by $\Delta\omega$. Figure 5-18 shows the result of this process. The top panel shows the Fourier transforms of the raw I and Q interferograms, which differ quite a bit since there's no reason for them to be the same. In contrast, the computed correlations in the bottom panel would be identical if not for the presence of noise and other imperfections in the system (likely residual saturation of the HEB). In fact, the two signals can even be averaged to improve the signal-to-noise ratio of the measurement, although the effect is marginal since the difference is only $\sqrt{2}$.

5.3.3 Bias dependence of comb formation

To show how the laser bias affects comb operation, plotted in Fig. 5-19(a) is the optical and RF power generated by a QCL comb versus its bias in the range over which the device lases. At the onset of strong RF emission, there is a drop in total optical power which corresponds to the onset of comb formation. Similarly, Fig. 5-19(b)

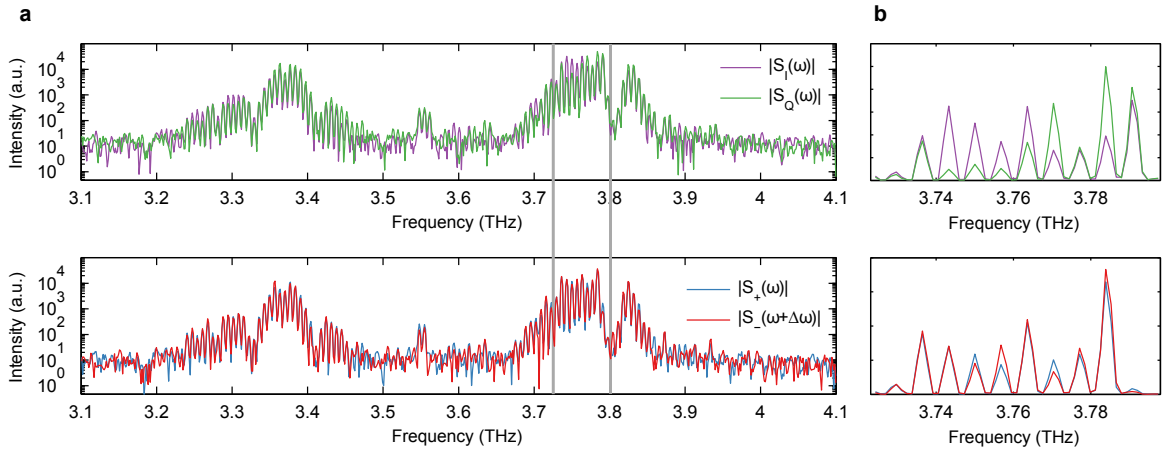


Figure 5-18: Consistency of SWIFT data. (a) Top panel: raw Fourier transforms of the I and Q interferograms of the data in Fig. 5-17. Bottom panel: calculated SWIFT coherences, with $S_{-}(\omega)$ shifted by the repetition rate of the laser. (b) The zoomed-in region is plotted on a linear scale.

shows the linewidth of the QCL's RF emission versus bias. Not all biases within the range of strong RF emission actually form frequency combs, as there are certain biases at which the beatnote linewidth abruptly increases to several MHz. Nonetheless, there are three distinct regions in which the beatnote linewidth is narrow and a stable comb is formed, denoted by Regions I, II, and III. In order to see whether these lasers can actually exploit the full gain bandwidth at a given laser bias, the SWIFT spectra obtained in the three regions of comb formation are compared with the gain spectra measured by terahertz time-domain spectroscopy. The result is plotted in Figure 5-19(c). In Region I, the lower-frequency lobe experiences more gain, and as a result the corresponding portion of the correlation spectrum dominates. In Region II, the two lobes of the gain and correlation spectra are approximately equal in strength. In Region III, the higher-frequency lobe is stronger and lases more strongly. This result demonstrates the effectiveness of the dispersion compensation technique, since it shows that frequency combs can be formed under various conditions with different gain profiles.

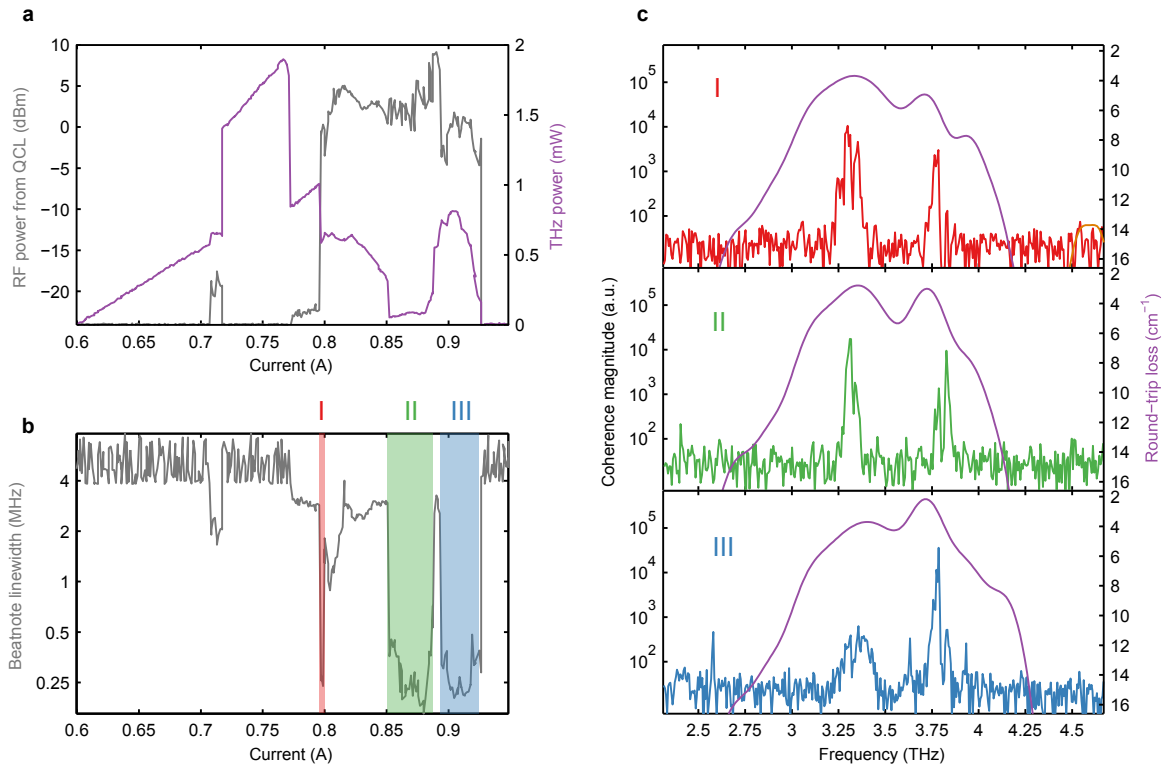


Figure 5-19: Bias dependence of beat-note and SWIFT spectra. (a) RF power measured from a bias-tee (amplified by 66 dB, with wiring losses of 20 dB) and calibrated terahertz power emitted by the QCL as a function of bias at 50 K, over the dynamic range of lasing operation. (b) Standard deviation of the RF signal emitted from the QCL as a function of bias, as measured with a spectrum analyser. The regions of stable comb formation are shaded and denoted I, II and III. (c) SWIFT coherence spectra (measured with the HEB) and gain spectra (measured with THz TDS) corresponding to each of the three regions.

5.3.4 SWIFTS for phase retrieval

Another feature of SWIFTS that has yet to be mentioned is its ability to resolve the phase difference between a pair of comb lines. Whereas normal FTS only measures $\langle |E(\omega)|^2 \rangle$, which is a strictly positive quantity, SWIFTS measures $\langle E^*(\omega)E(\omega \pm \Delta\omega)\rangle$, a quantity with phase. Ignoring the effects of apodization, the phase of SWIFTS is essentially $\angle S_+(\omega) = \angle E(\omega+\Delta\omega) - \angle E(\omega)$, that is the phase difference between adjacent lines of the comb. In fact, one can even say that this is essentially the principle on which SWIFTS operates: simultaneous measurement of the phase difference of

every comb pair.

If SWIFTS is used to retrieve all of the comb's phase differences, then in principle it can be used to retrieve the full phase of the electric field, by cumulative summing. This would make SWIFTS a rival to full-field pulse characterization techniques like FROG [105] and SPIDER [106]. In practice, integrating the phase is often impossible since any spectral gaps break the continuity. But unlike traditional pulse characterization techniques, which generally require a nonlinear element and are more suitable for optical pulses, SWIFTS only requires a fast linear detector. Moreover, SWIFTS can do one thing really well: measure the frequency-dependent group delay of the field coming from a comb. Since group delay is essentially $\tau_g \approx \frac{\Delta\phi}{\Delta\omega}$, this means that SWIFTS can be used to measure it modulo the cavity round-trip time.

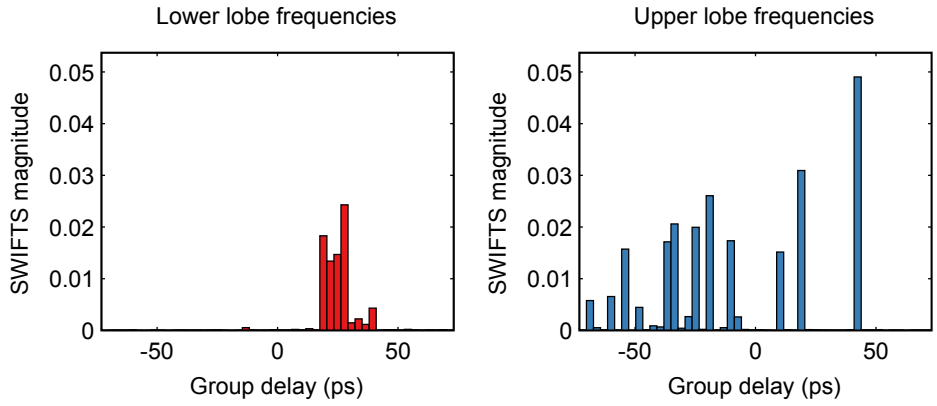


Figure 5-20: Group delay and coherence magnitude corresponding to the data shown in Figure 5-17. The data from the lower lobe of the gain spectrum is shown on the left; the data from the upper lobe is shown on the right.

Figure 5-20 shows the frequency-dependent group delay of each of the two portions of the gain spectrum, plotted against the corresponding magnitudes. All of the frequencies in the lower lobe essentially arrive at the same time, indicating that their behavior is pulse-like. In contrast, essentially none of the frequencies in the upper lobe—the stronger of the two lobes—arrived simultaneously. This indicates that no pulse is formed, and possibly that the output is strongly frequency-modulated. In general, it was found to be the case that the weaker of the two lobes could form pulses while the stronger of the two could not. Since four-wave mixing introduces a non-

trivial phase delay in frequency combs that resembles frequency modulation [107], this may indicate that the weaker lobe may not actually be forming a comb based on four-wave mixing. One possibility is that the stronger lobe actually introduces a weak microwave modulation in the cavity (that is, the beatnote), and that this microwave modulation effectively modulates the lower lobe.

One might wonder whether it is possible to extract more information about the time-domain behavior from SWIFTS. This is indeed the case, provided one is willing to live with the presence of noise. Imagine a hypothetical SWIFTS measurement in which it was determined that the spectrum consisted of two lobes separated by a null. Though SWIFTS could easily be used to determine the phase *difference* of each line in the comb, it could not be used to find the relative phase of the two lobes, since such a measurement is highly susceptible to noise. This is illustrated by Figure 5-21. In this plot, the two lobes have differing constant group delays—that is, they represent two pulses arriving at different times—but because the pulses overlap somewhat in the time domain the precise behavior of the time-dependent intensity is unknown, particularly in the region where the two pulses overlap. Nevertheless, there is clearly information there: the regions where the pulses do not overlap is very well-determined, despite the unknown phase. Therefore, it is useful to calculate the statistics of the measurement, shown in the last panel, which shows the mean and standard deviation of the time-dependent intensity.

In theory, any quantity that can be calculated from the field coefficients could be characterized in this way, including the electric field itself. In practice, one must restrict oneself to quantities which are not particularly phase-sensitive. (An attempt to characterize the electric field, for example, would give a result whose statistical fluctuations consume the entire measurement.) Fortunately, both the time-dependent intensity and the time-dependent frequency are amenable to this sort of measurement. If the field at frequency ω_n has a phasor E_n, then the field's positive frequency part is $E_+(t) = \sum_n E_n e^{j\omega_n t}$, and the time-dependent intensity and phase can be defined

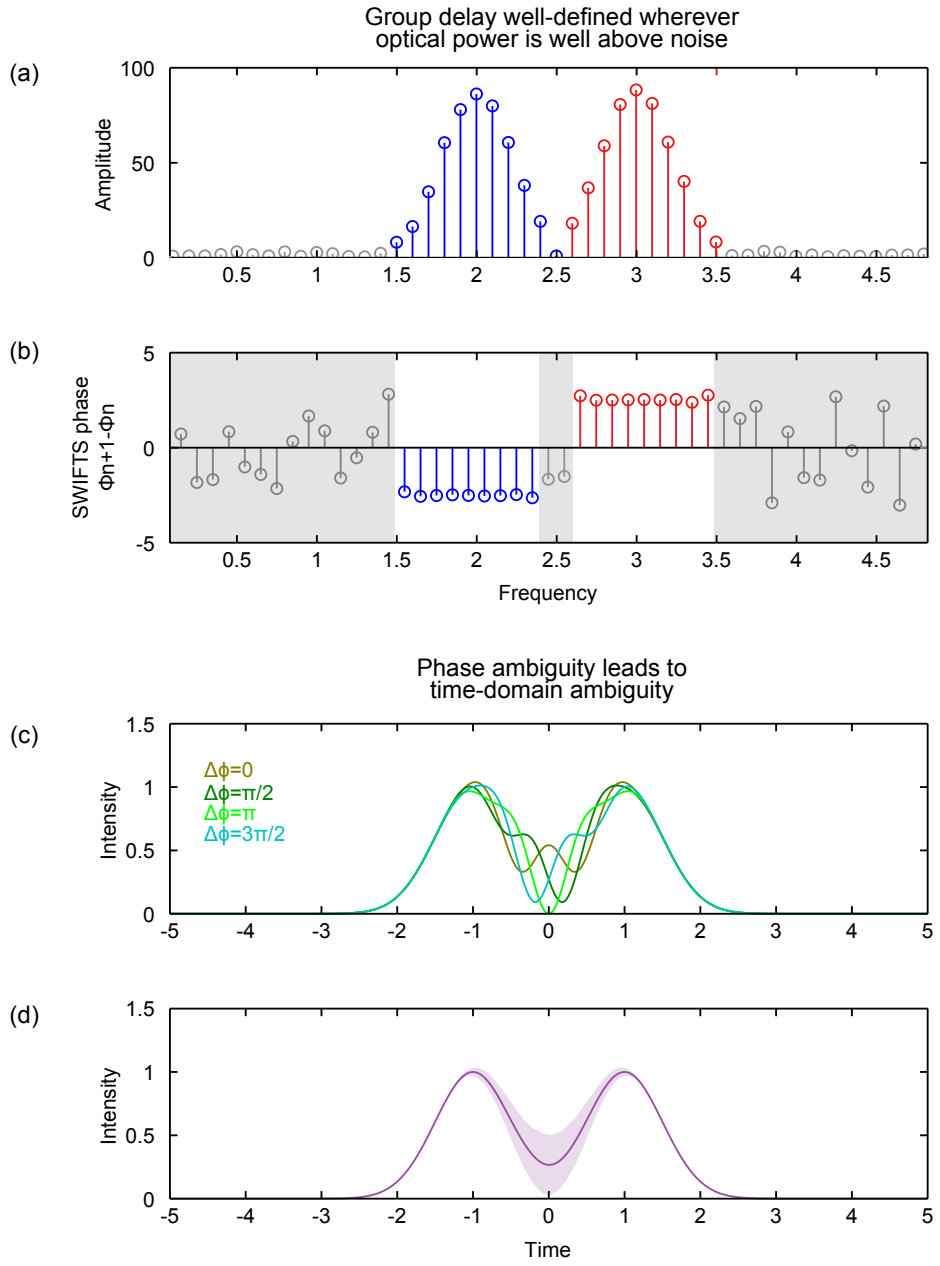


Figure 5-21: (a) and (b): Amplitude and phase of two lobes separated by a null. (c) Inferred time-dependent intensity assuming various relative phases. (d) Corresponding distribution of potential values. (The shaded region indicates a standard deviation.)

respectively as

$$I(t) \equiv E_+^* E_+ \tag{5.16}$$

$$f(t) \equiv \frac{1}{2\pi}\frac{d}{dt}\arg\left(E_+\right) = \frac{1}{2\pi|E_+|^2}\mathrm{Im}\left[E_+^*\frac{dE_+}{dt}\right], \tag{5.17}$$

where the identity $\frac{d}{dt}\arg(z) = \mathrm{Im}\left[z^* \frac{dz}{dt}\right] / |z|^2$ can be used to avoid discrete derivatives and phase unwrapping. Essentially, $I(t)$ represents the instantaneous intensity of the field and $f(t)$ represents the instantaneous carrier frequency. To make these quantities even more insensitive to phase fluctuations, it is often useful to filter with them with a reasonable integration time (say 10 ps) to reduce the standard deviation of the measurement without filtering out potentially informative time-domain features. This reduces sensitivity to phase error since both $I(t)$ and $f(t)$ are comprised wholly of terms which go like $E_+^* E_+$, giving rise to a double summation over frequency differences. Filtering the data removes frequency differences that are "far" from each other and are therefore more susceptible to statistical fluctations.

Another wrinkle is the issue of the type of randomness that is actually in the system. In Figure 5-21, the only random (unknown) variable was the relative phase between the two lobes. In reality, all frequencies possess some amount of noise which corrupt the SWIFTS measurement. Figure 5-22 shows a phasor diagram for two types of signals, a large one and a small one. The blue point represents the actual

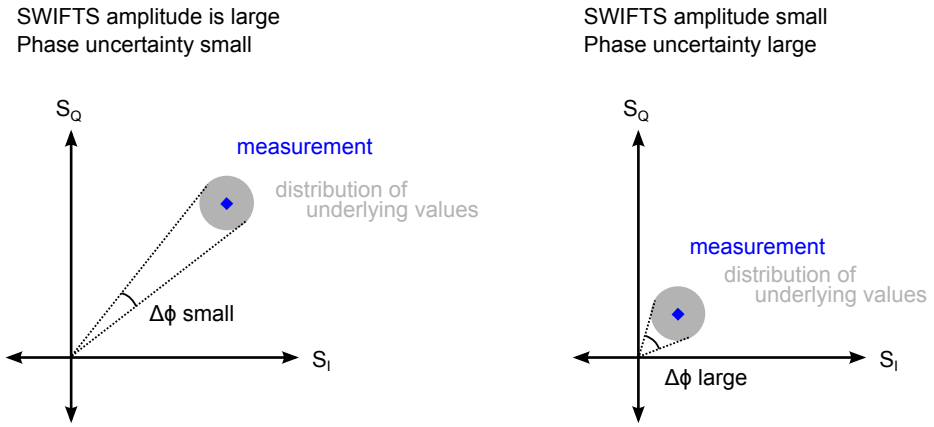


Figure 5-22: Phasor diagram of a large SWIFTS signal and a small one.

measurement, which has been corrupted by noise, whereas the gray region indicates the distribution of underlying values that could have produced this measurement. From this perspective, it is clear why the noise of some frequencies matter more than others. When the amplitude of the measured signal is large compared to the standard deviation of the noise, the uncertainty in phase is small. However, when the

amplitude of the measured signal is small, the uncertainty in phase is large. In fact, when the amplitude is down near the noise floor of the measurement, corresponding to a gap in the spectrum, the distribution of possible phases is effectively uniform over the interval $[0, 2\pi]$.

Strictly speaking, accounting for this uncertainty requires that the underlying probability distribution of the SWIFTS signal be known at each frequency. While this could be accomplished by doing many measurements, it is easier to note that when the noise of each point of the interferogram is uncorrelated and the interferogram is large, the central limit theorem dictates that the noise in each channel of the FFT is Gaussian (regardless of the type of noise in the interferogram).[3] Therefore the noise can be modeled using a two-dimensional Gaussian whose standard deviation is determined by the noise floor.

Analytically calculating the statistics of the time-dependent quantities is challenging, and so Monte Carlo methods were used instead. The following procedure was used:

1. Use a region of the SWIFTS spectrum known to be noisy to find the standard deviation of each quadrature of the measurement, σ_I and σ_Q. In principle one should measure these as a function of frequency, but in practice it is usually easier to assume that the noise is white and that the "worst case" value is used.

2. Take the observed SWIFTS data, $S_{I/Q}(\omega)$, and subtract from each frequency a different value drawn from the normal distribution. Use it to find a possible value of the "true" SWIFTS signal uncorrupted by noise, and denote it by $S_+^{(i)}(\omega)$.

3. For each iteration, use $S_+^{(i)}(\omega)$ and $S_0(\omega)$ to find the phasors associated with the nth comb lines. Denote these by $E_n^{(i)}$. Use (5.16) and (5.17) to determine the time-dependent intensity and frequency, and filter the results in the time domain.

[3] If x_n represents the interferogram samples, the real part of the DTFT at frequency ω is $X(\omega) = \sum_n x_n cos(\omega n)$. This is a weighted sum, and so the central limit theorem applies.

4. Repeat steps 2 and 3 for a large number of samples. Calculate the expectation values $\langle I(t) \rangle$ and $\langle f(t) \rangle$, as well as the confidence level for each measurement (which is also time-dependent).

Figure 5-23 shows the result of this process for the data previously shown (at 0.9 A). First, examine the intensity versus time. As was claimed using the group

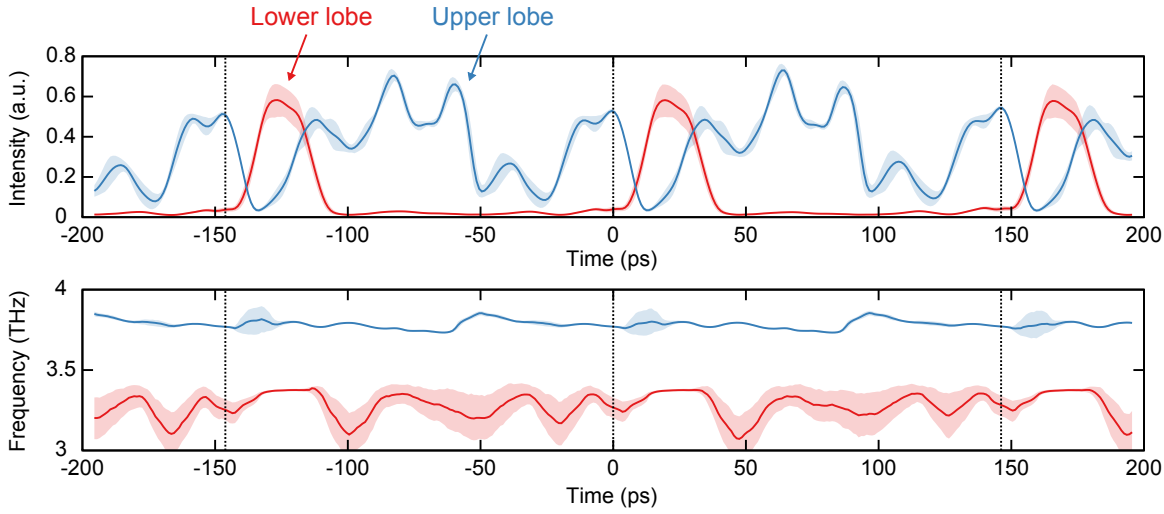


Figure 5-23: Time-dependent intensity and frequency inferred from the previous SWIFTS data, filtered with a 10 ps window. Shaded regions indicate a standard deviation, and dotted lines indicate a single period (round-trip time).

delay plot, the lower (red) lobe of the gain spectrum produces a time-domain pulse that arrives at t=20 ps. Even though there is some uncertainty in the height of the pulse near its peak, it is clear that phase noise doesn't change the fact that it is a pulse. The upper (blue) lobe of the gain spectrum, meanwhile, lases at practically all times during the laser's period *except* for the time at which the lower lobe is lasing. In other words, the lower lobe can only lase by virtue of the upper lobe shutting off, and is essentially a result of temporal hole burning. Shifting focus to the frequency versus time, an important thing to note is that the standard deviation of the frequency measurement grows large whenever the intensity of each lobe is small, which is of course expected since the signal-to-noise ratio at these times is low. The red lobe is essentially completely flat in the region where its SNR is high, meaning

that the pulse is not chirped and is nearly transform-limited. In contrast, the blue lobe is strongly frequency-modulated and has a sawtooth-like profile in the time-domain. This is indicative of a self-frequency modulation that is a signature of laser microcombs [107].

This information is extremely useful for characterizing the different regimes of comb operation, perhaps even more so than the SWIFTS measurement directly. Figure 5-24 shows the bias dependence of $I(t)$ measured using SWIFTS alongside the different regimes of comb operation. The three regions of comb operation have strikingly different time-domain behavior. Region I is dominated by the red lobe, and exhibits a complicated multi-pulse emission. The blue lobe emits only when the red lobe transiently turns off. In region II, the two signals have similar average powers, but the red lobe's power is concentrated in narrow time-domain pulses while the blue lases nearly continuously. It is a regime of passive mode-locking, albeit for only part of the spectrum. In region III, the blue lobe dominates and the red lobe lases only when it shuts off. Once again, a signature of temporal hole burning is evident.

5.4 Absolute coherence of comb

Though most of this chapter has focused on mutual coherence, the absolute linewidth of the comb will also matter for applications like dual-comb spectroscopy. The reason for this is that to have a unique mapping between the optical domain and RF domain requires that all of the beatnotes between the two combs fit within half the repetition rate. If N is the number of laser lines, Δ is the detuning between repetition rates, and f_r is the repetition rate, then this condition can be succinctly stated as $N\Delta < f_r/2$ [87]. For combs based on solid-state mode-locked lasers, the repetition rate is small and N is large, so linewidths on the order of hundreds of Hz are required. In contrast, a comb with 2 THz of bandwidth and a repetition rate of 6.8 GHz only needs to fit 294 lines into a 3.4 GHz span, resulting in a required separation of 11.5 MHz.

To measure the comb's absolute coherence, it is beat with a narrow-linewidth distributed-feedback (DFB) laser. Since the comb is already known to be mutually

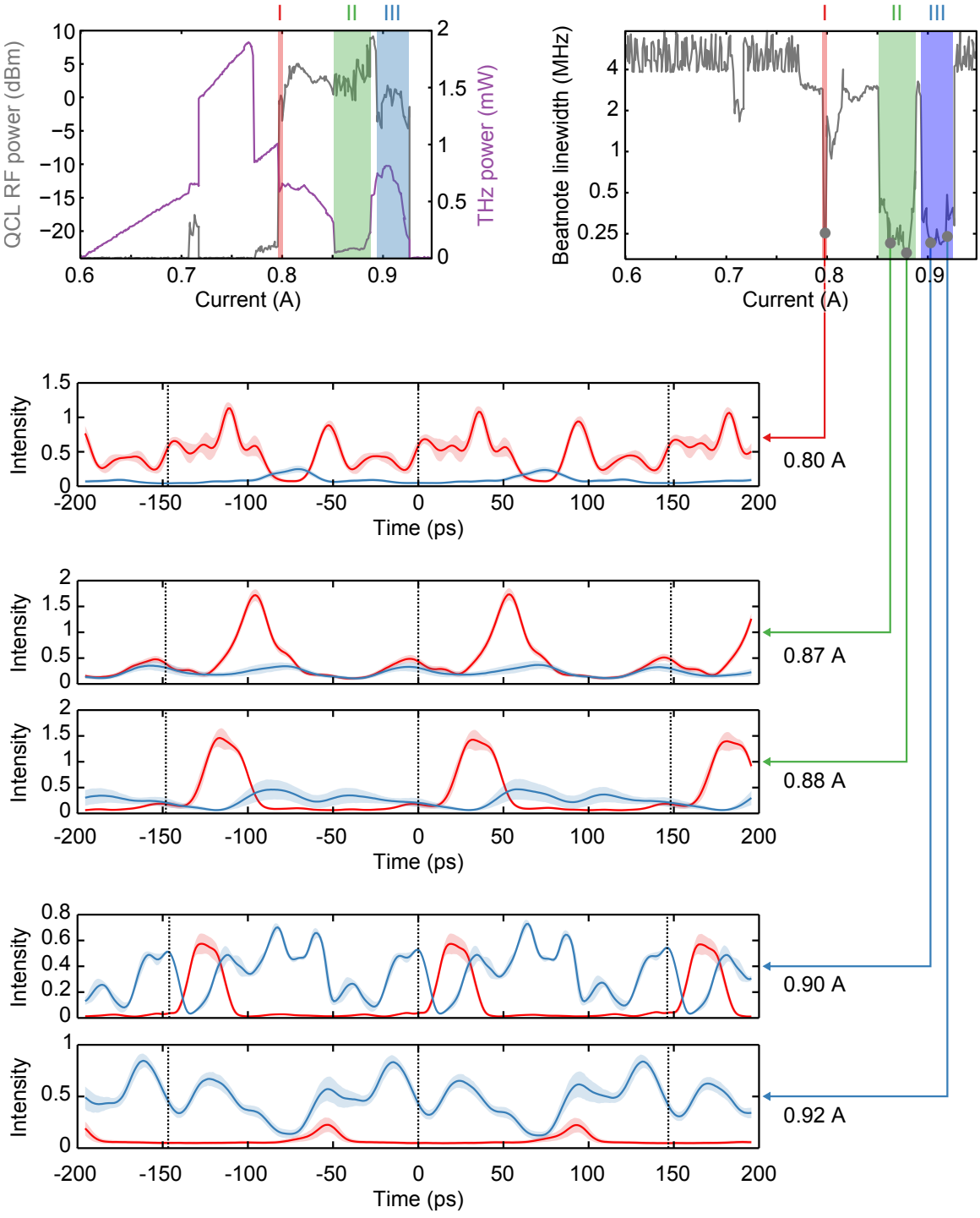


Figure 5-24: Bias dependence of time-dependent intensities. Once again, red indicates lower lobe frequencies while blue indicates lower lobe frequencies.

coherent to Hz levels thanks to SWIFTS, measuring the linewidth of a single line will essentially measure the linewidth of all the lines. A third-order DFB [108, 109] was constructed using metal-metal waveguides fabricated from the same gain medium as the comb, FL183S. It was 21 µm wide, 450 µm long, and consisted of 15 periods of a corrugation that lased at 3.85 THz, with about a milliwatt of power. The DFB was operated in continuous-wave mode in a compact Stirling cryocooler at 47 K. For this measurement, the DFB laser was shined *directly* onto the facet of the comb laser, and the heterodyne beating generated by the intracavity mixing of the DFB laser and the comb laser was examined. Though fast detectors like HEBs or Schottky mixers can be used to measure similar properties, this measurement also demonstrates one of the salient features of this particular comb: the ability to detect the frequency of terahertz lasers without any additional components (modulo the repetition rate). The setup is shown in Figure 5-25. The RF beatnotes were collected from the comb

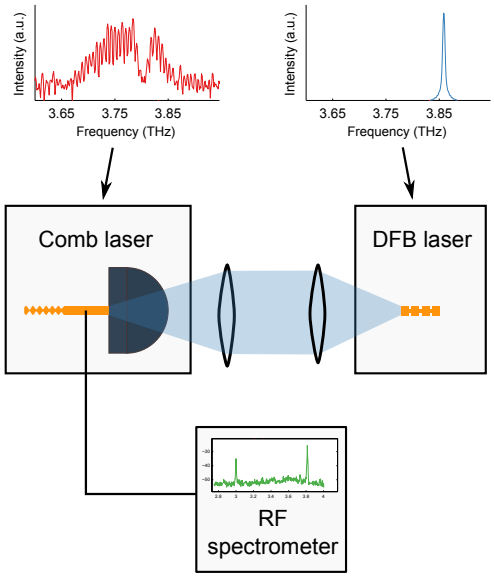


Figure 5-25: Setup used for measuring the heterodyne beating between a comb laser and a DFB laser.

laser using a bias tee and amplified with a low-noise amplifier providing 50 dB of gain.

Figure 5-26(a) shows the spectra of the two lasers involved: a single-mode DFB laser at approximately 3.85 THz, and the SWIFT spectrum of a comb biased near

0.9 A. Figure 5-26(b) shows the beatnotes generated by a comb device over a large

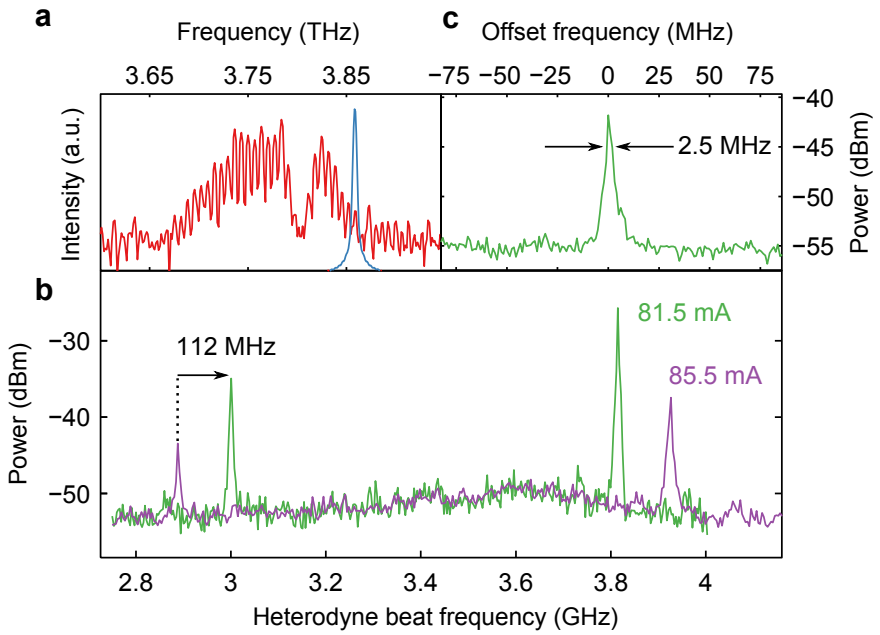


Figure 5-26: Heterodyne beatnote between single-mode laser and comb. (a) Coherence spectrum of comb at 0.9 A, spectrum of DFB laser. (b) Heterodyne beating of free-running comb laser at 0.885 A with free-running DFB laser at various biases. (c) Zoomed-in view of one of the lines, showing a convolved FWHM of 2.5 MHz.

frequency span, measured with a resolution bandwidth of 3 MHz. At a given bias, there are two lines present below the repetition rate, corresponding to beating of the DFB laser with the two nearest adjacent comb lines. Their frequencies therefore add up to the comb's repetition rate. Moreover, when the DFB's frequency is shifted by tuning its bias, the two lines are observed to move in opposite directions by the same amount, the laser's tuning coefficient (28 MHz/mA in this case, the same as what was measured by a collaborator in Ref. [110]). Even though the DFB laser is operating near the edge of the comb's bandwidth and only a few percent of the DFB's power can be coupled into the metal-metal waveguide, it is still possible to measure beatnotes with signal-to-noise ratios exceeding 20 dB. Figure 5-26(c) shows one of these beatnotes over a smaller frequency span. Because the DFB's frequency and comb's offset frequency are both unstabilized, the measured full-width half maximum (FWHM) linewidth of 2.5 MHz represents a convolution between their respective

lineshapes. Assuming they are identical Gaussians, it is therefore possible to estimate an absolute comb linewidth of 1.8 MHz. Previous characterization of similar DFB lasers has measured free-running linewidths between 1 and 3 MHz [111], showing that the comb's offset frequency is as stable as a single-mode laser. Even in the worst case scenario, this linewidth is more than suitable for dual-comb spectroscopy at these wavelengths, owing to the large ratio of the laser's repetition rate to its frequency. If this laser was used to perform dual-comb spectroscopy, its linewidth could perform a measurement with a resolution equivalent to a FT spectrometer with 60 meters of delay.

Chapter 6

Conclusions and future work

In summary, this thesis used terahertz quantum cascade lasers as broadband sources, samples, and detectors. By applying dispersion compensation techniques first developed at visible wavelengths to the THz QCL, I was able to make compact frequency combs based on THz QCLs that passively generated broadband radiation. Such combs cover a frequency range of almost 500 GHz with more than 70 lines at 3.5 THz, and represent nearly an almost order of magnitude improvement over the 10 modes obtained by active mode-locking. The comb's bandwidth covers 14% of its center frequency—the highest fractional bandwidth of integrated semiconductor frequency combs to date—suggesting that similar techniques can be used to improve laser microcombs at other wavelengths, including the mid-infrared.

In developing a means for characterizing these combs I demonstrated SWIFTS, a coherent detection scheme that can be used in combination with FTS to quantitatively measure the performance of such lasers and to characterize the efficacy of comb formation. I have also demonstrated that by utilizing intracavity mixing these lasers can be used to compactly measure the frequency of single-mode lasers without the need for a high-speed terahertz detector or an external solid state laser. There is much future work to be done. At present, the gain medium' double-peaked shape causes the laser's energy to be split across two lobes and limits its general applicability, as the region between the two lobes experiences too much loss and cannot be accessed without a high-dynamic range detector. In addition, the weak interaction

between the two lobes leads to a complicated bias dependence of the comb's properties and reduces its robustness. Fortunately, both of these issues are straightforward to resolve, as continuing development of heterogeneous gain media will lead to flatter, broader comb spectra. Octave-spanning gain spectra were recently achieved in THz QCLs [75], and once combs are made from them it may even be possible to stabilize the comb's absolute frequency without any additional components, since the intracavity mixing process demonstrated in this thesis should also be capable of generating f-2f beating. Such combs could be used to form compact solid-state dual-comb spectrometers, which by utilizing intracavity mixing may even be able to operate without external detectors. These devices could potentially even act as complete terahertz spectrometers on a chip, as shown in Figure 6-1.

This thesis also used broadband terahertz time-domain spectroscopy to analyze the behavior of THz QCLs. By using QCLs as independent photoconductive switches, the usual limitations imposed by optical coupling are circumvented, and properties of the laser previously inaccessible can be directly observed. These properties include the gain and absorption of the laser gain medium above and below threshold, anticrossings in the system, the populations of the laser's subbands, and properties of the waveguide like its loss and dispersion. The temperature performance of various lasing schemes were analyzed, and this information was used to identify a additional transitions contributing to the laser's performance at high temperatures. Knowledge of these properties were used to guide frequency comb design, and were also used to inform simulations for designing better lasers including genetic algorithm-designed structures.

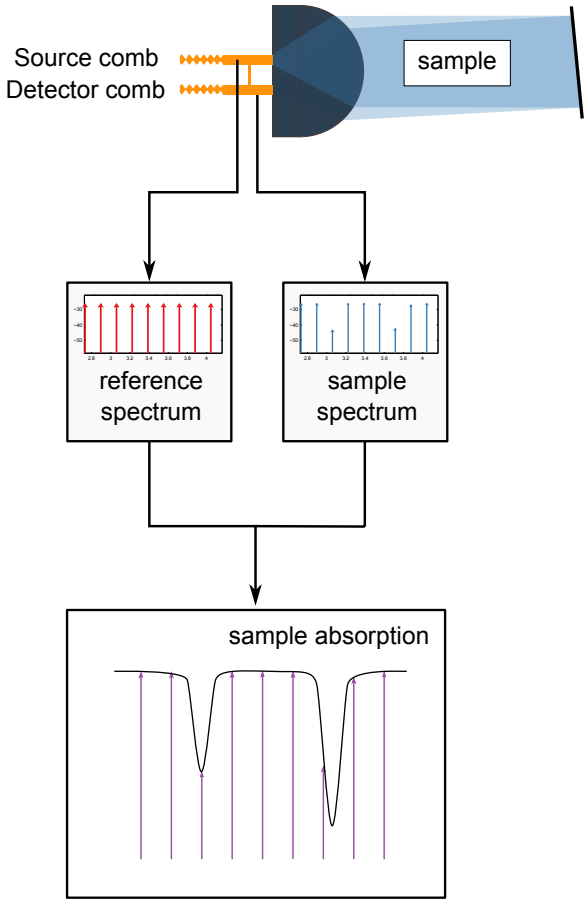


Figure 6-1: Ideal dual comb spectrometer. The sources and detector combs are co-integrated, f-2f stabilization is done internally, and a weak coupling element allows the source to detect the spectroscopic reference. Everything is electronic.

Appendix A

Periodic density matrix formalism

Density matrices are a useful tool for analyzing the dynamics of QCLs, on account of their ability to coherently model the effects of both transport and the electromagnetic interaction. This appendix reviews the basic principles of density matrix analysis and develops a periodic density matrix formalism particularly well-suited for describing periodic systems like QCLs. This formalism is completely egalitarian in the sense that it requires no human to decide where a module starts and where it ends, making it especially suitable for algorithmic optimization (for example, using genetic algorithms).

A.1 Density matrices of a finite system

A.1.1 Basic definitions

In theory, the dynamics of an electron existing in a pure state $|\psi\rangle$ are completely determined by Schrödinger's equation, $\hat{H}|\psi\rangle = i\hbar\frac{\partial}{\partial t}|\psi\rangle$. In reality, one nevers knows the precise state of the wavefunction or even the exact form of the Hamiltonian. For example, in QCLs layer thickness fluctuations will stochastically modify the Hamiltonian in a way that depends on the local environment. Density matrices can account for this. Suppose that an electron is only known to be in one of many possible states of a large ensemble, and that the probability of an electron being in state $|\psi_i\rangle$ is p_i.

The density matrix associated with this electron is defined to be

$$\hat{\rho} \equiv \sum_i p_i |\psi_i\rangle\langle\psi_i| = \Big\langle |\psi\rangle\langle\psi| \Big\rangle. \tag{A.1}$$

Here, $\hat{\rho}$ is an operator. The "matrix" part of "density matrix" comes from projecting this operator into a finite orthogonal basis $\{|n\rangle\}$, i.e. by defininining $\rho_{nm} = \langle n|\hat{\rho}|m\rangle$. When written this way the defining relation becomes

$$\rho_{nm} = \Big\langle \langle n|\psi\rangle\langle\psi|m\rangle \Big\rangle = \langle c_n c_m^* \rangle \,, \tag{A.2}$$

where the values of $\{c_n\}$ represent the coefficients of the basis states. In other words, density matrices represent the wavefunction's second-order statistics. The on-diagonal elements, for which $n = m$ and $\rho_{nn} = \langle |c_n|^2 \rangle$, are referred to as *populations* since they are strictly positive numbers that represent how occupied a state is. The off-diagonal elements, for which $n \neq m$ and $\rho_{nn} = \langle c_n c_m^* \rangle$, are referred to as *coherences* since they represent how phase-coherent states n and m are.[1]

One of the most important aspects of quantum mechanics is the calculation of expectation values of an operator $\hat{A}$, and this can of course be done using density matrices. Doing this requires that the expectation value of each wavefunction in the ensemble be averaged:

$$\begin{aligned} \langle \hat{A} \rangle &= \sum_i p_i \langle \psi_i | \hat{A} | \psi_i \rangle = \sum_i p_i \langle \psi_i | \left(\sum_{nm} A_{nm} |n\rangle\langle m| \right) |\psi_i\rangle \\ &= \sum_{inm} A_{nm} p_i \langle m|\psi_i\rangle\langle\psi_i|n\rangle = \sum_{nm} A_{nm}\rho_{mn} = \sum_n (A\rho)_{nn} = \mathrm{Tr}(A\rho). \end{aligned} \tag{A.3}$$

Calculating expectation values therefore only requires a calculation of the trace of an operator with the density matrix. Note that because the trace operation is invariant under a unitary change of basis, expectation values remain basis-independent (as expected). This is assuming that the density matrix has been normalized to $\mathrm{Tr}(\rho) =$

[1] The same analysis that was used to motivate the definition of optical coherence applies here. Two states are said to be completely coherent if $\rho_{nm}/\sqrt{\rho_{nn}\rho_{mm}} = \langle c_n c_m^* \rangle \,/\, |c_n|\,|c_m| = 1$.

$\sum_i p_i = 1$; other normalizations can be used but they must be divided out when expectation values are calculated. Some other properties of density matrices are summarized below, although they are not derived:

- $\hat{\rho}$ is Hermitian, and so $\rho^\dagger = \rho$ and $\rho_{nm} = \rho^*_{mn}$.
- $\mathrm{Tr}(\rho^2) = 1$ when a state is pure (i.e., comprised of a single wavefunction).
- $\mathrm{Tr}(\rho^2) < 1$ when a state is mixed (i.e., comprised of multiple wavefunctions).
- $|\rho_{nm}|^2 \leq \rho_{nn}\rho_{mm}$ for $n \neq m$, with equality holding only for pure states. Therefore pure states are the most coherent.
- If a wavefunction had a representation in some basis v, then the density matrix corresponding to a pure state of that wavefunction is simply an outer product, i.e. $\rho^{(v)} = vv^\dagger$.
- The matrix inner product of two density matrices is defined by $\langle \rho_a, \rho_b \rangle \equiv \mathrm{Tr}(\rho_a^\dagger \rho_b) = \mathrm{Tr}(\rho_a \rho_b)$.
- If the wavefunctions n and m are orthogonal, then the density matrices of the corresponding pure states will be orthogonal in the sense that $\rho^{(n)}\rho^{(m)} = \rho^{(n)}\delta_{nm}$ and $\langle \rho^{(n)}, \rho^{(m)} \rangle = \delta_{nm}$.

A.1.2 Time-evolution

Next, the dynamics of density matrices are considered. First, consider how they evolve as a result of Schrödinger's equation, $\hat{H}\,|\psi\rangle = i\hbar\frac{\partial}{\partial t}\,|\psi\rangle$, the so-called coherent part of the evolution:

$$\begin{aligned}\frac{\partial \hat{\rho}}{\partial t} &= \frac{\partial}{\partial t}\left(|\psi\rangle\langle\psi|\right) = \left\langle \frac{\partial|\psi\rangle}{\partial t}\langle\psi| + |\psi\rangle\frac{\partial\langle\psi|}{\partial t} \right\rangle \\ &= \left\langle \frac{1}{i\hbar}\hat{H}|\psi\rangle\langle\psi| - \frac{1}{i\hbar}|\psi\rangle\langle\psi|\hat{H} \right\rangle = \frac{1}{i\hbar}\left(\hat{H}\hat{\rho} - \hat{\rho}\hat{H}\right) = \frac{1}{i\hbar}\left[\hat{H}, \hat{\rho}\right]. \end{aligned} \tag{A.4}$$

Here, brackets represent commutators. In matrix form, this is often written simply as

$$\dot{\rho} = \frac{1}{i\hbar}\left[H, \rho\right]. \tag{A.5}$$

This result is essentially the density matrix equivalent to Schrödinger's equation and is often referred to as Liouville's equation. The consequence of this is that even if all of the individual wavefunctions comprising the original ensemble were unknown, as long as the density matrix is known one can use it to fully describe the system's observables and time evolution. Therefore, the density matrix is a complete description of reality.

Still, this is basically just a warmed-over version of Schrödinger's equation. Where density matrices really shine is in their ability to handle the effects of incoherent processes, including scattering and dephasing. A common approach for modeling transport through semiconductor structures is using rate equations, with scattering rates determined by Fermi's Golden Rule. However, this approach fails whenever wavefunction coherence is important to transport. By including these semiclassical scattering events in Liouville's equation, both incoherent and coherent processes can be considered simultaneously. First, consider the impact of scattering from state i to state f. If this occurs with a time constant of τ_{fi}, then one would expect that the population (on-diagonal) element of the density matrix will decay with this rate, i.e. $\rho_{ii}(t) = \langle |c_{ii}(t)|^2 \rangle \sim \langle |c_{ii}(0)|^2 \rangle e^{-t/\tau_{fi}}$, and that the population of the final state will increase according to this schedule. In other words, $\dot{\rho}_{ii} = -\rho_{ii}/\tau_{fi}$ and $\dot{\rho}_{ff} = \rho_{ii}/\tau_{fi}$. Likewise, it is reasonable to require that the coherences associated with state i decay with half this rate since coherence is associated with fields and not intensity, i.e. $\rho_{ij}(t) \sim \langle c_i(0) c_j^*(0) \rangle e^{-t/2\tau_{fi}}$ and $\dot{\rho}_{ij} = -\rho_{ij}/2\tau_{fi}$. All of these conditions can be compactly specified as

$$\dot{\rho}_{nm} = -\frac{\rho_{nm}}{2\tau_{fi}}\left(\delta_{ni} + \delta_{mi}\right) + \frac{\rho_{ii}}{\tau_{fi}}\delta_{nf}\delta_{mf}, \tag{A.6}$$

where the first term takes care of the decay of state i while the second takes care of the transfer to state f. Note that the decay of the coherences is typically referred to as dephasing, since it corresponds to a decay in the phase stability of the two

components. Equation (A.6) only includes a single scattering process; including all scattering processes $\{\tau_{nm}\}$ results in

$$\begin{aligned}\dot{\rho}_{ij} &= \sum_{nm} -\frac{\rho_{ij}}{2\tau_{nm}}\left(\delta_{im}+\delta_{jm}\right)+\frac{\rho_{mm}}{\tau_{nm}}\delta_{in}\delta_{jn} \\ &= -\rho_{ij}\left(\frac{1}{2\tau_i}+\frac{1}{2\tau_j}\right)+\delta_{ij}\sum_n \frac{\rho_{nn}}{\tau_{jn}},\end{aligned} \tag{A.7}$$

where $\frac{1}{\tau_i} \equiv \sum_n \frac{1}{\tau_{ni}}$ is the total lifetime of the ith state. Note that this ansatz allows for the presence of terms that scatter states into themselves; these processes scramble the wavefunction's phase but do not transfer population and are referred to as pure dephasing. Typically, an additional phenomenological dephasing term $1/T_2^*$ is added to (A.7) to include the effects of scattering processes that are not modeled. With this modification, the scattering term becomes

$$\dot{\rho}_{ij} = -\rho_{ij}\left(\frac{1}{2\tau_i}+\frac{1}{2\tau_j}+\frac{1}{T_2^*}\right)+\delta_{ij}\sum_n \frac{\rho_{nn}}{\tau_{in}} \equiv (\Gamma\rho)_{ij}. \tag{A.8}$$

Combining the coherent and incoherent terms, the dynamical equation for the density matrix is given by

$$\boxed{\dot{\rho} = \frac{1}{i\hbar}[H,\rho]+\Gamma\rho \equiv \mathcal{L}\rho}, \tag{A.9}$$

where $\mathcal{L}$ is called the Liouville superoperator.

A.1.3 Superoperators

How can Liouville's equation (A.9) be solved to find physically meaningful quantities? The key observation is that despite the fact that ρ is a matrix, the operator $\mathcal{L}$ is still linear in ρ. Therefore, all of the usual machinery of vector spaces apply to ρ, and if it is reshaped into a column vector then the action of the Liouville superoperator is just the action of a matrix. If the system had N states then the density matrix would be $N \times N$, the vectorized density matrix would be $N^2 \times 1$, and the vectorized superoperator would be $N^2 \times N^2$. A superoperator requires four indices to describe it: two for the source term and two for the destination term. For example, (A.9)

could be written as

$$\dot{\rho}_{ij} = \sum_{kl} \mathcal{L}_{ij,kl}\rho_{kl}. \tag{A.10}$$

Once everything has been vectorized, Liouville's equation can be solved in equilibrium by setting the time derivative to zero and finding the ρ_0 that satisfies $\mathcal{L}\rho_0 = 0$ and $\mathrm{Tr}(\rho_0) = 1$. This solution is guaranteed to exist and to be unique as long as some dissipation is present in the system. Though it is possible to find the explicit form of the superoperator $\mathcal{L}$ using brute force, it is easier to do with elementary operations. To do this, it is often convenient to make several definitions:

- The Kronecker product, which forms the outer product of two matrices and therefore forms a superoperator, is defined by $(A \otimes B)_{ij,kl} \equiv A_{ik}B_{jl}$.

- The Hadamard product, $A \odot B$, is simple entrywise multiplication of two matrices.

- The vectorization operator, $\mathrm{vec}(A)$, is the operation that turns a matrix into a vector; the precise ordering of the elements is determined by convention (e.g., column-normal order). Some basic properties:

  - $\mathrm{vec}(AB) = (I \otimes A)\,\mathrm{vec}(B) = (B^{\dagger *} \otimes I)\,\mathrm{vec}(A)$, where I is the identity matrix.

  - $\mathrm{Tr}(A^\dagger B) = \mathrm{vec}(A)^\dagger\,\mathrm{vec}(B)$. The matrix and vector inner products coincide.

  - $\mathrm{vec}(A \odot B) = \mathrm{vec}(A) \odot \mathrm{vec}(B)$

For example, the coherent part of the superoperator can be expressed using Kronecker products, since

$$\begin{aligned}\mathrm{vec}\left(\frac{1}{i\hbar}[H,\rho]\right) &= \frac{1}{i\hbar}\,\mathrm{vec}(H\rho - \rho H) \\ &= \frac{1}{i\hbar}\left(I \otimes H - H^{\dagger *} \otimes I\right)\,\mathrm{vec}(\rho)\end{aligned}$$

Other terms can be expressed similarly. The advantage of this approach is that the superoperators can be directly constructed using simple expressions instead of more complicated index-dependent ones.

A.1.4 Optical susceptibility of density matrices

Perhaps the most important quantity of interest when modeling QCLs is not the electron distribution itself or even the current density, but rather how the QCL affects light that is propagating through it. For this, we must perform a first-principles calculation of the susceptibility. The approach taken here is to perturb the system with an optical field and to use first-order perturbation theory to see how the density matrix responds. Assume that in the absence of a perturbation the system has a Hamiltonian H_0 and a density matrix ρ_0, with $\dot{\rho}_0 = \mathcal{L}\rho_0$. If the system is weakly perturbed with a Hamiltonian H', then the first order correction to the density matrix ρ' must obey Liouville's equation:

$$\begin{aligned}
\dot{\rho}_0 + \dot{\rho}' &= \frac{1}{i\hbar}\left[H_0 + H', \rho_0 + \rho'\right] + \Gamma(\rho_0 + \rho') \\
&= \underbrace{\frac{1}{i\hbar}[H_0, \rho_0] + \Gamma\rho_0}_{\mathcal{L}\rho_0} + \underbrace{\frac{1}{i\hbar}[H_0, \rho'] + \Gamma\rho'}_{\mathcal{L}\rho'} \\
&\qquad + \frac{1}{i\hbar}[H', \rho_0] + \underbrace{\frac{1}{i\hbar}[H', \rho']}_{\text{2nd-order}} \\
\dot{\rho}' &= \mathcal{L}\rho' + \frac{1}{i\hbar}[H', \rho_0].
\end{aligned} \tag{A.11}$$

This result can be used to calculate linear susceptibilities; nonlinear susceptibilities require that higher order perturbations be calculated. To calculate the response to an optical field, assume a plane wave of the form $\vec{E}(t) = E(\omega)e^{i\omega t}\hat{z}$ is impinged. The corresponding vector potential in the Coulomb gauge will be given by $\vec{E} = -\frac{\partial \vec{A}}{\partial t}$, or $\vec{A}(t) = -\frac{1}{i\omega}E(\omega)e^{i\omega t}\hat{z}$. For electrons with a mass m_0 and a charge $-e$ the

corresponding perturbation potential is

$$\begin{aligned} H'(t) &= \frac{e}{m_0}\vec{A}\cdot\vec{p} \\ &= i\frac{e}{m_0\omega}E(\omega)e^{i\omega t}p_z, \end{aligned} \tag{A.12}$$

where p_z is the component of the electron's momentum in the growth direction. This form is usually converted into a dipole moment using the position operator, but as will be discussed subsequently this is inconvenient since the position operator is not well-defined in periodic systems. Assume that the matrix elements of p_z have been evaluated in the same basis as the density matrix. If the unperturbed system is in equilibrium, the density matrix ρ_0 will be constant in time, and by (A.11) $\rho'(t)$ will have the same frequency components as $H'(t)$. Therefore, we can write without loss of generality that $\rho'(t) = \rho'(\omega)e^{i\omega t}$. Inserting the perturbation Hamiltonian into the first-order result, we find that

$$\begin{aligned} (i\omega - \mathcal{L})\,\rho'(\omega) &= \frac{e}{m_0\hbar\omega}[p_z,\rho_0]E(\omega) \\ \rho'(\omega) &= E(\omega)\frac{e}{m_0\hbar\omega}\,(i\omega-\mathcal{L})^{-1}\,[p_z,\rho_0]. \end{aligned} \tag{A.13}$$

Note that since $\mathcal{L}$ is a superoperator, $(i\omega-\mathcal{L})^{-1}$ is a matrix inversion that acts on the commutator $[p_z,\rho_0]$, a matrix.

To find the permittivity of the QCL gain medium, ε_r, assume that the total electric displacement field is a sum of the component arising from the intersubband contribution and from the material (which has a refractive index n):

$$\begin{aligned} \varepsilon_0\varepsilon_r\frac{\partial\vec{E}}{\partial t} &= \vec{J}_\rho + \varepsilon_0 n^2\frac{\partial\vec{E}}{\partial t} \\ \varepsilon_r(\omega)E(\omega) &= \frac{1}{i\omega\varepsilon_0}J_\rho(\omega) + n^2E(\omega) \\ \varepsilon_r(\omega) &= n^2 + \frac{1}{i\omega\varepsilon_0}\frac{J_\rho(\omega)}{E(\omega)}. \end{aligned} \tag{A.14}$$

Therefore, we need to find the semiclassical current density associated with the in-

tersubband transitions. Current density is the product of charge density and charge velocity, and in QCLs with a doping level of N_d is essentially $\vec{J} = -eN_d\langle\vec{v}\rangle = -\frac{eN_d}{m_0}\langle\vec{p}\rangle$. The portion of this current density directed in the growth direction at a frequency ω is just

$$\begin{aligned} J_z(\omega) &= -\frac{eN_d}{m_0}\mathrm{Tr}\left(p_z\rho'(\omega)\right) \\ &= -\frac{e^2N_d}{m_0^2\hbar\omega}\mathrm{Tr}\left(p_z\left(i\omega-\mathcal{L}\right)^{-1}\left[p_z,\rho_0\right]\right)E(\omega), \end{aligned}$$

leading to the final result that

$$\boxed{\varepsilon_r(\omega) = n^2 + i\frac{e^2N_d}{\varepsilon_0 m_0^2\hbar\omega^2}\mathrm{Tr}\left(p_z\left(i\omega-\mathcal{L}\right)^{-1}\left[p_z,\rho_0\right]\right)}.^2 \tag{A.15}$$

This expression can be used to calculate the effective index and optical gain for essentially any electronic system, using $n(\omega) = \mathrm{Re}(\sqrt{\varepsilon_r(\omega)})$ and $g(\omega) = 2\omega/c\ \mathrm{Im}(\sqrt{\varepsilon_r(\omega)})$. It essentially requires only the superoperator, the equilibrium density matrix, and the momentum matrix. Note that because dephasing built into the superoperator already broadens the transition linewidths, no phenomenological broadening is necessary beyond that already built into the superoperator. The cost of this generality is that attributing features to any one optical transition is difficult, as gain essentially comes from the whole density matrix at once. While this makes it appealing from a computational standpoint, it is unappealing from a pedagogical standpoint.

A.2 Periodic density matrices

The analysis in the previous section was completely general, applying to any quantum system. In principle, it should be applied to all of the states of a QCL. Of course, this is not computationally practical; in reality the system should be reduced to just a few modules. Traditionally, this is done by cutting the module at the injector

[2]Evaluating this trace requires that $(i\omega-\mathcal{L})^{-1}$ be calculated for each value of ω. As $\mathcal{L}$ can be quite large, it is computationally unwieldly. Instead, it is usually faster to diagonalize $\mathcal{L}$ first.

barrier and requiring that electrons be injected through this channel [34]. However, this approach is not without problems. For example, it is not always clear *where* the injector barrier is for an arbitrary system, or indeed whether one exists at all. Ideally, algorithmic design of QCL bandstructures would be unencumbered by traditional design paradigms, and should treat every layer equally.

For this, a theory of periodic density matrices was developed. This formalism is essentially an extension of the common Kazarinov-Suris formalism, and can be used to write down and solve for the dynamical equations of a system like the QCL without cutting the module at an arbitrary location or elevating one layer above the rest. To start, imagine writing down the full density matrix of a large system with N_{mod} modules and N states per module. Call this full matrix ρ_f. If the density matrix were truly periodic, one would expect that the elements of the density matrix repeat down the diagonal, as shown below:

$$\rho_f = \begin{pmatrix} \ddots & & & & & & & & \\ & \rho_{-1} & & & & & & & \\ & \rho_0 & \rho_{-1} & & & & & & \\ & \rho_1 & \rho_0 & \boxed{\rho_{-1}} & & & & & \\ & & \rho_1 & \boxed{\rho_0} & \rho_{-1} & & & & \\ & & & \boxed{\rho_1} & \rho_0 & \rho_{-1} & & & \\ & & & & \rho_1 & \rho_0 & & & \\ & & & & & \rho_1 & \ddots & & \\ & & & & & & & & \end{pmatrix} \tag{A.16}$$

In this expression, each ρ_i is considered to be an $N \times N$ matrix and is referred to as a *block*; blocks represent the coherence between all the states of a module and all the states of another module. If $\rho_{ij}^{(f)}$ represents the block at position ij of ρ_f, then we would write $\rho_{ij}^{(f)} = \rho_{i-j}$. Note that the off-diagonal blocks are not independent of each other, since Hermiticity guarantees that $\rho_{-i} = \rho_i^\dagger$. The highlighted submatrix of ρ_f is referred to as a *tile* and is essentially a column array of blocks. It contains all of the information about the system, and will be denoted by ρ_t.

In the spirit of tight-binding, it is assumed to be the case that the states of a module can only interact with the states of nearby modules. $\mathcal{N}$ will be used to

represent this distance ($\mathcal{N} = 1$ for nearest-neighbor coupling). Usually it is more convenient to deal with $M \equiv 2\mathcal{N} + 1$, the number of modules that each module interacts with. Because we will not care about coherences that may develop between modules that are far apart, blocks any further than $\mathcal{N}$ off the diagonal are neglected. It is for this reason that the tile matrix ρ_t is an $M \times 1$ array (i.e., an $MN \times N$ matrix).

Though the tile matrix is a complete description of the density matrix, it is not square and can be inconvenient for calculations. To form a square matrix, let $\mathcal{S}$ denote the circular shift operator that replicates and shifts a column vector M times. Then the *periodicized* version of the tile matrix is defined as $\rho_\infty \equiv S\rho_t$. For example, for M=3:

$$\rho_\infty \equiv \begin{pmatrix} \rho_0 & \rho_{-1} & \rho_1 \\ \rho_1 & \rho_0 & \rho_{-1} \\ \rho_{-1} & \rho_1 & \rho_0 \end{pmatrix} \tag{A.17}$$

Another way of formulating this definition is by writing $\rho^{(\infty)}_{ij} \equiv \rho_{[i-j]}$, where brackets are used to denote modulo M. In this section, all rows and column indices will run from $-\mathcal{N}$ to $\mathcal{N}$, and so the modulo operation is assumed to map back into this region.[3] The key advantage of using periodicized operators is that nearly all of the usual machinery of density matrices applies to them directly, with only a few modifications. Even though they are redundant, they can also be reduced to tile operators quite easily as ρ_t is just the center column of ρ_∞, that is $\rho_t = \rho^{(\infty)}_{\bullet 0}$.[4]

A.2.1 Periodicization of operators

Other quantum mechanical operators can be periodicized in the same way as the density matrix, since an operator A_f will have a form similar to ρ_f. However, there is one very important difference. Whereas ρ_f is completely periodic, operators that depend explicitly on position will usually be shifted on their diagonal from one module to the next. For example, the position operator will be shifted by the length of a

[3] Explicitly, $[x] \equiv \mathrm{mod}(x + \mathcal{N}, M) - \mathcal{N}$.

[4] Bullets appearing in indices are used here to denote rows and columns of matrices, for example $A_{\bullet j}$ is the jth column of A.

module, and the Hamiltonian will be shifted by the voltage drop per module. A general operator A_f must therefore be written as

$$A_{ij}^{(f)} = A_{i-j} + \lambda_A I \delta_{ij} j \tag{A.18}$$

where λ_A is the module-to-module shift and I is the $N \times N$ identity matrix. Similarly, the periodicized operator A_∞ must be defined in such a way that the shift is incorporated, that is

$$A_{ij}^{(\infty)} \equiv A_{[i-j]} + \lambda_A I \delta_{ij} j \tag{A.19}$$

$$A_\infty = \begin{pmatrix} A_0 - \lambda_A I & A_{-1} & A_1 \\ A_1 & A_0 & A_{-1} \\ A_{-1} & A_1 & A_0 + \lambda_A I \end{pmatrix} \quad \text{(for M=3).} \tag{A.20}$$

Of course, the operator A_∞ is not really periodic, and so this means that one must take great care when assigning properties to it. In this analysis, expectation values are only calculated for operators which lack a shift, like the momentum operator p_z. It is for this reason the optical susceptibility was left in terms of the momentum operator rather than converting it a dipole moment.

What are the expectation values of an operator in terms of periodic density matrices and tiles? Assume that the block representing a module's interaction with itself, ρ_0, has been normalized to $\mathrm{Tr}(\rho_0) = 1$. In other words, the probability of finding an electron in a module is 1. Then the expectation value of the full density matrix is simply the usual expectation value, normalized to the number of modules:

$$\langle A \rangle = \frac{1}{N_{mod}} \mathrm{Tr}(A_f \rho_f). \tag{A.21}$$

For shift-free operators, this result can be expressed in terms of blocks as follows

(keeping in mind that ρ_i is a matrix, not a number):

$$\begin{aligned}\langle A\rangle &= \frac{1}{N_{mod}}\mathrm{Tr}(A_f\rho_f) = \frac{1}{N_{mod}}\sum_{n=1}^{N_{mod}}\mathrm{Tr}\left((A_f\rho_f)_{nn}\right) = \mathrm{Tr}\left((A_f\rho_f)_{00}\right)\\ &= \mathrm{Tr}\left(\sum_{k=-\mathcal{N}}^{\mathcal{N}} A^{(f)}_{0k}\rho^{(f)}_{k0}\right) = \sum_k \mathrm{Tr}(A_{-k}\rho_k)\\ &= \sum_k \mathrm{Tr}(A_k^\dagger\rho_k) = \mathrm{Tr}(A_t^\dagger\rho_t). \end{aligned} \tag{A.22}$$

This allows for the calculation of expectation values directly from tiles or from blocks.[5] It can also be calculated from periodic operators by performing a similar calculation, in which case one finds that $\mathrm{Tr}\left(A_\infty\rho_\infty\right) = M\sum_k \mathrm{Tr}(A_k^\dagger\rho_k)$, or

$$\langle A\rangle = \frac{1}{M}\mathrm{Tr}\left(A_\infty\rho_\infty\right). \tag{A.23}$$

The factor of M is of course due to the fact that periodic operators contain M copies of a tile. The following table summarizes these results for each representation of the density matrix, in decreasing order of size:

Type	Symbol	Block representation	Size	$\langle A\rangle$
full	ρ_f	$\rho^{(f)}_{ij} = \rho_{i-j}$	$N_{mod}N \times N_{mod}N$	$\frac{1}{N_{mod}}\mathrm{Tr}(A_f\rho_f)$
periodic	ρ_∞	$\rho^{(\infty)}_{ij} = \rho_{[i-j]}$	$MN \times MN$	$\frac{1}{M}\mathrm{Tr}\left(A_\infty\rho_\infty\right)$
tile	ρ_t	$\rho^{(t)}_i = \rho_i$	$MN \times N$	$\mathrm{Tr}(A_t^\dagger\rho_t)$
blocks	$\{\rho_i\}$	$\rho_i = \rho_i$	M $(N \times N)$'s	$\sum_k \mathrm{Tr}(A_k^\dagger\rho_k)$

A.2.2 Periodicization of superoperators

To find the dynamics of a system requires that Liouville's equation be translated into the language of periodic density matrices. However, this is not as straightforward as the calculation of expectation values, since special care must be taken to ensure that the time-evolution of periodic density matrices maintains periodicity. Assume that the superoperator of the finite system, $\mathcal{L}_f$, has been found. What is the corresponding

[5] Since A_t is not a square matrix, $A_t^\dagger \neq A_t$.

superoperator for tiles, $\mathcal{L}_t$? Note that in contrast to the full superoperator which takes four indices to describe each block, the tile superoperator only needs two indices since it tiles are column arrays. Consider finding $\mathcal{L}^{(t)}_{-11}$, which represents the contribution of ρ_1 to $\dot{\rho}_{-1}$. Since ρ_1 and ρ_{-1} appear many times in ρ_f, there will be many ways ρ_1 can contribute to $\dot{\rho}_{-1}$:

$$\rho_f = \begin{array}{c} \begin{array}{ccccc} -2 & -1 & 0 & 1 & 2 \end{array} \\ \left(\begin{array}{ccccc} \ddots & & & & \\ \rho_{-1} & & & & \\ \rho_0 & \rho_{-1} & & & \\ \rho_1 & \rho_0 & \rho_{-1} & & \\ & \rho_1 & \rho_0 & \rho_{-1} & \\ & & \rho_1 & \rho_0 & \rho_{-1} \\ & & & \rho_1 & \rho_0 \\ & & & & \rho_1 \\ & & & & \end{array} \right) \begin{array}{c} \\ -3 \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array} \qquad \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \ddots \end{array} \tag{A.24}$$

In general, a "source" block ρ_j appears at location $(j, 0)$ in the full density matrix, whereas the possible "destination" blocks appear at $(i + t, t)$, where t is any integer. Therefore, we can write that the superoperator associated with ρ_t is given by

$$\boxed{\mathcal{L}^{(t)}_{ij} = \sum_{t=-\infty}^{\infty} \mathcal{L}^{(f)}_{(i+t)t,j0}} \tag{A.25}$$

Note that as before, whole blocks are being indexed, meaning that $\mathcal{L}^{(t)}_{ij}$ is not a number, it is a superoperator that acts on $N \times N$ matrices.

To find the superoperator for periodic matrices, $\mathcal{L}_\infty$, the process is similar, but instead of considering individual blocks the entire periodic matrix needs to be tiled across ρ_f and a macro-periodicity of M is assumed:

$$\rho_f = \begin{array}{c} \begin{array}{ccccc} -2 & -1 & 0 & 1 & 2 \end{array} \\ \left(\begin{array}{ccccc} \ddots & & & & \\ \rho_{01} & & & & \\ \rho_{11} & \rho_{1\bar{1}} & & & \\ \rho_{\bar{1}1} & \rho_{\bar{1}\bar{1}} & \rho_{\bar{1}0} & & \\ & \rho_{0\bar{1}} & \rho_{00} & \rho_{01} & \\ & & \rho_{10} & \rho_{11} & \rho_{1\bar{1}} \\ & & & \rho_{\bar{1}1} & \rho_{\bar{1}\bar{1}} \\ & & & & \rho_{0\bar{1}} \quad \ddots \end{array} \right) \end{array} \begin{array}{c} -3 \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \tag{A.26}$$

A source block ρ_{kl} appears at location $(l-[l-k], l)$ in the full density matrix, whereas the destination block ρ_{ij} appears at locations $(j-[j-i]+Mt, j+Mt)$. Therefore, the superoperator associated with ρ_∞ is given by

$$\mathcal{L}^{(\infty)}_{ij,kl} = \sum_{t=-\infty}^{\infty} \mathcal{L}^{(f)}_{(j-[j-i]+Mt)(j+Mt),(l-[l-k])l}$$

Shifting t to make $j-[j-i] = [j-[j-i]] = i$,

$$\boxed{\mathcal{L}^{(\infty)}_{ij,kl} = \sum_{t=-\infty}^{\infty} \mathcal{L}^{(f)}_{(i+Mt)(i-[i-j]+Mt),(l-[l-k])l}} \tag{A.27}$$

Even though this expression looks fairly complicated, as a result of the nearest neighbor approximation it tends to simplify dramatically.

The superoperators associated with dephasing and scattering are rather easy to periodicize, but the one associated with coherent transport—that is, arising from the Hamiltonian—is less straightforward. Nevertheless, the result itself is exceptionally clean, and so it is worth doing as an example. To find it, note that the time-evolution of the full density matrix is $i\hbar\dot{\rho}_f = \sum_{kl} H^{(f)}_{ik}\delta_{lj}\rho_{kl} - \rho_{kl}\delta_{ik}H^{(f)}_{lj}$. Inserting this form into (A.27), we find that

$$\begin{aligned} i\hbar\dot{\rho}^{(\infty)}_{ij} = \sum_{kl}\sum_{t=-\infty}^{\infty} & H^{(f)}_{(i+Mt)(l-[l-k])}\delta_{l(i-[i-j]+Mt)}\rho_{kl} \\ & - \rho^{(\infty)}_{kl}\delta_{(i+Mt)(l-[l-k])}H^{(f)}_{l(i-[i-j]+Mt)} \end{aligned}$$

By modular arithmetic, $l = i - [i-j] + Mt$ implies $l = j$, and $l - [l-k] = i + Mt$ implies $k = i$. Using this to remove the delta functions and changing the l's in the second term to k's,

$$i\hbar\dot{\rho}^{(\infty)}_{ij} = \sum_k \sum_{t=-\infty}^{\infty} H^{(f)}_{(i+Mt)(j-[j-k])}\rho^{(\infty)}_{kj} - \rho^{(\infty)}_{ik} H^{(f)}_{k(i-[i-j]+Mt)}$$

Next, the shift-periodicity of the Hamiltonian, $H^{(f)}_{nm} = H_{n-m} + \lambda_H I\delta_{nm}n$, is used. However, note that because of the nearest-neighbor form of the Hamiltonian, $H_{x+Mt} = 0$ for H too far from the diagonal (i.e., H outside the mod range). As a consequence, $\sum_t H_{x+Mt} = H_{[x]}$, and the $(i + Mt) - (j - [j-k])$ term becomes $[i-k]$ while the $k - (i - [i-j] + Mt)$ term becomes $[k-j]$:

$$\begin{aligned} i\hbar\dot{\rho}^{(\infty)}_{ij} = \sum_k &(H_{[i-k]} + \lambda_H\delta_{[i-k]}(j - [j-k]))\rho^{(\infty)}_{kj} \\ &- \rho^{(\infty)}_{ik}(H_{[k-j]} + \lambda_H\delta_{[k-j]}k) \end{aligned}$$

Since $H_{[n-m]} = H^{(\infty)}_{nm} - \lambda_H I\delta_{nm}n$,

$$\begin{aligned} i\hbar\dot{\rho}^{(\infty)}_{ij} &= \sum_k \left(H^{(\infty)}_{ik} + \lambda_H\delta_{ik}(j-k-[j-k])\right)\rho^{(\infty)}_{kj} - \rho^{(\infty)}_{ik}H^{(\infty)}_{kj} \\ &= \sum_k H^{(\infty)}_{ik}\rho^{(\infty)}_{kj} - \rho^{(\infty)}_{ik}H^{(\infty)}_{kj} + \lambda_H\delta_{ik}\left([i-j] - (i-j)\right)\rho^{(\infty)}_{ij} \\ &= \left(H_\infty\rho_\infty - \rho_\infty H_\infty\right)_{ij} + \lambda_H\left([i-j] - (i-j)\right)\rho^{(\infty)}_{ij} \end{aligned}$$

The first term is just a commutator and the second just an entrywise multiplication, so we at last arrive at our final result:

$$\boxed{i\hbar\dot{\rho}_\infty = [H_\infty, \rho_\infty] + \lambda_H C_\infty \odot \rho_\infty} \tag{A.28}$$

where $C_{ij} \equiv [i-j] - (i-j)$ and $C^{(\infty)}_{ij} \equiv C_{ij}\mathbb{1}_N$ (an $M \times M$ and $MN \times MN$ matrix, respectively). In other words, the periodic density matrix evolves nearly identically to the aperiodic one, with the exception of a correction term that incorporates the

effect of the shift. Note that C has an extremely simple form; e.g. for M=3 and 5

$$C_3 = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ -3 & 0 & 0 \end{pmatrix} \qquad C_5 = \begin{pmatrix} 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ -5 & 0 & 0 & 0 & 0 \\ -5 & -5 & 0 & 0 & 0 \end{pmatrix}$$

These matrices only affect the off-diagonal elements of Liouville's equation, and essentially correct for the error introduced by the shift in H_∞. A similar result holds for the tile and for the block matrices:

$$i\hbar\dot{\rho}_i = \sum_j H_{[i-j]}\rho_j - \rho_j H_{[i-j]} + \lambda_H k\delta_{ik}\rho_i \tag{A.29}$$

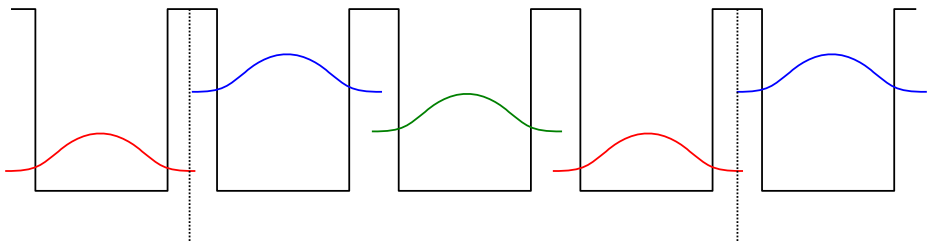
$$i\hbar\dot{\rho}_t = (\mathcal{S}H_t)\,\rho_t - (\mathcal{S}\rho_t)\,H_t + \lambda_H D_t \odot \rho_t \tag{A.30}$$

where $\mathcal{S}$ is the circular shift operator, $D_k \equiv k$ and $D_k^{(t)} \equiv D_k \mathbb{1}_N$.

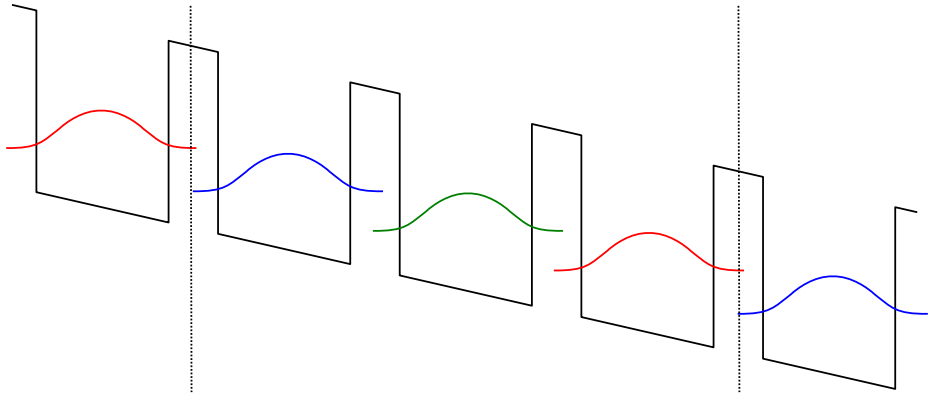
As an example, consider the basic biased superlattice with one state per module and nearest-neighbor coupling ($N = 1$ and $M = 3$). If the coupling between a well and its neighbor is Δ, and the bias per module is λ, then the periodic Hamiltonian would be

$$H_\infty = \begin{pmatrix} -\lambda & \Delta^* & \Delta \\ \Delta & 0 & \Delta^* \\ \Delta^* & \Delta & \lambda \end{pmatrix} \tag{A.31}$$

As it stands, this Hamiltonian actually reflects the Hamlitonian of the wrong infinite system, effectively this one:



instead of this one:



It is impossible to formulate Schrodinger's equation in such a way that the biased system can ever be modeled, short of explicitly including each and every state. However, the same is not true of Liouville's equation:

$$i\hbar\dot{\rho}_\infty = \underbrace{\begin{pmatrix} 0 & -\lambda\rho_{-1} & -2\lambda\rho_1 \\ \lambda\rho_1 & 0 & -\lambda\rho_{-1} \\ 2\lambda\rho_{-1} & \lambda\rho_1 & 0 \end{pmatrix}}_{[H_\infty,\rho_\infty]} + \lambda \underbrace{\begin{pmatrix} 0 & 0 & 3\rho_1 \\ 0 & 0 & 0 \\ -3\rho_{-1} & 0 & 0 \end{pmatrix}}_{C_\infty \odot \rho_\infty} = \begin{pmatrix} 0 & -\lambda\rho_{-1} & \lambda\rho_1 \\ \lambda\rho_1 & 0 & -\lambda\rho_{-1} \\ -\lambda\rho_{-1} & \lambda\rho_1 & 0 \end{pmatrix}$$

The key point here is that the uncorrected result is not periodic and reflects the dynamics of the wrong system, whereas the corrected result is fully periodic. Moreover, the factor of 3λ essentially accounts for the incorrect positioning of the periodic extension.

A.3 Basis for scattering calculations

Up until now, the choice of basis used for constructing the density matrix has not been mentioned. This is because in principle, any orthogonal basis should be suitable for the preceding analysis. Still, there is one major exception: the calculation of scattering rates. Scattering should always be calculated in the energy eigenbasis since Fermi's Golden Rule requires energy eigenstates as inputs. There are two problems with this:

1. The effect of dephasing on the choice of basis is not always apparent. The usual approach is to restrict the band structure simulation to one module or another, effectively solving for the non-injector eigenstates in a delocal basis and solving for the injector eigenstates in a local basis. However, as previously mentioned this presupposes that an injector barrier even exists. It is also mathematically discontinuous, making it a less than ideal choice.

2. In a periodic system, it is not clear whether a thing such as energy eigenstates even exist. The corrected Liouville's equation is for density matrices, not wavefunctions. Moreover, at first glance it is not immediately obvious how wavefunctions even fit into this formalism.

A single solution can solve both issues: an orthogonal basis that minimizes the time evolution of the density matrix, $||\dot{\rho}||^2$. But first, it is necessary to talk about the choice of computational basis.

A.3.1 Computational basis

The choice of an appropriate basis can be tremendously helpful for analyzing different structures. A useful concept for analyzing band structures algorithmically is the concept of a layer function $L(z)$, a function of position that ticks up every time a new layer is encountered:

$$L(z) = k \qquad \text{for} z \in \text{layer } k. \tag{A.32}$$

One can also define similarly a module function $M(z)$, which ticks up every time a new QCL module is encountered. The reason these operators find a use is that they can be used to construct wavepackets which nicely localize wavefunctions in a layer or a module, without destroying the interactions between adjacent states as solving Schrodinger's equation with infinite barriers would. To do this, suppose that the wavefunctions of a finite system under flat-band conditions, $\psi_n(z)$, have already been found. The matrix elements of the layer operator can be calculated using $L_{nm} = \int_{-\infty}^{\infty} \psi_n^*(z) L(z) \psi_m(z)$, and the resulting matrix L can be diagonalized in

terms of its eigenvalues. The resulting eigenvectors are essentially the finite basis' best representation of the layer function, and in analogy to LCAO the resulting eigenstates minimize the standard deviation of the layer operator. (In concept, the layer-localized basis functions are similar to Wannier functions.) Figure A-1 shows the local basis functions calculated for FL183S, along with the usual wavefunctions.

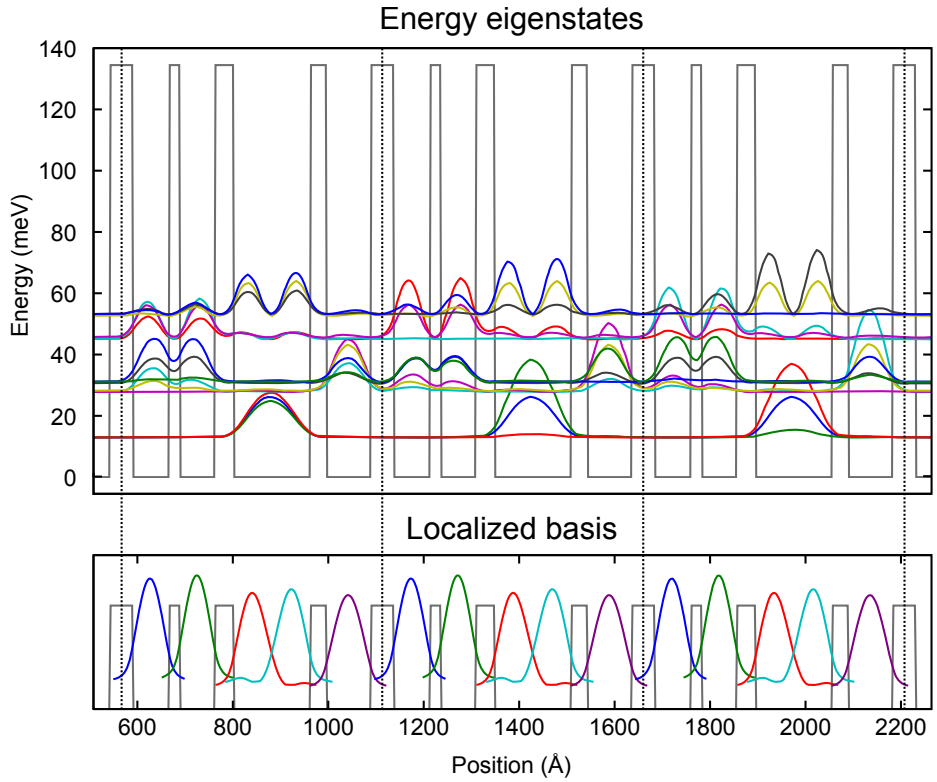


Figure A-1: Band diagram of FL183S gain medium in the absence of bias, in both the eigenbasis and in the layer-localized basis.

Even when large numbers of states per module are included ($N > 10$), the layer-localized basis states agree with the "intuitive" notion of QCLs as a series of weakly-coupled wells. An advantage of the layer-localized basis is that it can be used to implement pure dephasing due to layer thickness variations directly. The eigenvalues of the layer operator correspond to the layer each basis function is localized to, and so one expects that the dephasing rate between two layer states due to interface roughness is proportional to the difference in eigenvalues. Pure dephasing was usually implemented in this basis.

A.3.2 Quasi-eigenstates

To find the appropriate basis for scattering calculations, it is worth considering the meaning of an energy eigenstate. Because an electron in an energy eigenstate ψ_n will stay in that state indefinitely, the corresponding density matrix associated it is completely stationary ($\dot{\rho_n} = 0$). In a system without dissipation, there are N such states, and as a result the Liouville superoperator has a null space with dimensionality N. In contrast, a system with any dephasing and scattering between all levels will only permit a unique $\dot{\rho}_0 = 0$—the equilbrium value—which is not even a pure state. Accordingly, finding "eigenstates" of a system with dephasing is impossible.

In spite of this, if the requirement that $\dot{\rho_n} = 0$ is relaxed, it is possible to define *quasi-eigenstates* that are similar to energy eigenstates. Imagine that a collection of wavefunctions $\{\psi_j\}$ have coefficients V_{ij} in some basis (i.e., ψ_j is represented by the column vector $V_{\bullet j}$). The pure state density matrix that corresponds to this function is $\rho_j = V_{\bullet j} V_{\bullet j}^\dagger$, and the corresponding time evolution is $\dot{\rho}_j = \mathcal{L} V_{\bullet j} V_{\bullet j}^\dagger$. One can then define a cost function which represents the amount of time evolution experienced by V:

$$\epsilon(V) \equiv \sum_j \left|\left| \mathcal{L} V_{\bullet j} V_{\bullet j}^\dagger \right|\right|^2 \tag{A.33}$$

This cost function is fourth-order in V, and can be readily minimized using numerical optimization techniques since its gradient can be analytically determined. The main issue is that eigenstates which are correctly-defined should be orthogonal to each other, meaning that V needs to be optimized under the constraint that it be unitary. This was done by projecting the gradient onto a Riemannian space and using rotation matrices to ensure that unitarity was preserved [113]. A secondary issue that pertains especially to algorithmic design of QCLs is that such an optimization can be computationally expensive if the superoperator $\mathcal{L}$ is left in its matrix form. Having analytical expressions for the time-evolution of the density matrix (e.g., (A.28)) without explicitly constructing superoperators is often extremely helpful in this regard.

For typical values of the phenomenological pure dephasing parameter and of the

injector anti-crossing, the quasi-eigenstates closely resemble the wavefunctions normally obtained by cutting the module at the injector barrier, only without a human necessary to make this judgment. Once a suitable basis has been found, the expectation value of the Hamiltonian is used as the quasi-eigenstates' "energy," and scattering rates are calculated using Fermi's Golden Rule. The superoperator associated with this scattering, Γ, is constructed, and is rotated back[6] into the computational basis using

$$\Gamma \rightarrow (V^* \otimes V)\,\Gamma\,(V^* \otimes V)^\dagger\,. \tag{A.34}$$

A.3.3 Periodic eigenstates

The notion of quasi-eigenstates extends quite naturally to periodic density matrices: as there are N basis functions per module one just needs to find N quasi-eigenstates. Still, there are some caveats. The first point of note is that a wavefunction in the periodic formalism must be allowed to extend over M modules to allow for nearest-neighbor coupling. The prototypical example would be a laser whose upper laser level is aligned to its injector. In the absence of dephasing the energy eigenstates will be symmetric and anti-symmetric linear combinations of the basis states, even though those basis states are in separate modules. Essentially, this means that the coefficient matrix V should be the same size as a tile matrix ($MN \times N$), and can be written in terms of blocks in the same way (e.g., for M=3):

$$V_t = \begin{pmatrix} V_{-1} \\ V_0 \\ V_1 \end{pmatrix} \tag{A.35}$$

[6]When rotating into a new basis V, a vector w becomes $V^\dagger w$, a matrix A becomes $V^\dagger A V$, and a superoperator $\mathcal{L}$ becomes $(V^* \otimes V)^\dagger\,\mathcal{L}\,(V^* \otimes V)$.

As always, one can define a periodicized version of V_t by considering its circularly-shifted tiling, $V_\infty = \mathcal{S}V_t$:

$$V_\infty = \begin{pmatrix} V_0 & V_{-1} & V_1 \\ V_1 & V_0 & V_{-1} \\ V_{-1} & V_1 & V_0 \end{pmatrix} \tag{A.36}$$

Another caveat lies in the unitarity constraint. We should not merely require that V_t be unitary (i.e., that $V_t^\dagger V_t = I$), we should also require that V_∞ be unitary (i.e., that $(\mathcal{S}V_t)^\dagger(\mathcal{S}V_t) = I$). In other words, the columns of V_t must also be orthogonal to shifted versions of themselves. Essentially, what this means is that not only do wavefunctions within a module need to be orthogonal to each other, wavefunctions in different modules also need to be orthogonal to each other. Fortunately, all of the properties that pertain to unitary matrices also pertain to periodic unitary matrices, and so this constraint is not difficult to satisfy.

The last caveat lies in the construction of a density matrix associated with a given wavefunction. Normally, the pure state associated with a state ψ would be constructed from an outer product of its coefficient vector, $vv^\dagger$. However, implicit in the periodic density matrix formalism is the notion that even *if* we knew that the system was in state ψ we would still lack information about *which* module it was in. Therefore, the best we can hope to do is to construct a mixed state consisting of a superposition of the same wavefunction tiled over all modules. Letting $\mathcal{S}_i$ be a modified circular shift operator that shifts an $MN \times MN$ matrix down and to the right by iN,[7] the state associated with v should be defined by

$$\rho_\infty^{(v)} = \sum_i \mathcal{S}_i \left(vv^\dagger\right). \tag{A.37}$$

In fact, it is this choice of construction that ensures that $\rho_{j\infty}$ is periodic as required. It is also interesting to note that even though this is not a pure state *per se*, if states v and w were shift-orthogonal then it would be the case that $\rho_\infty^{(v)}\rho_\infty^{(w)} = \rho_\infty^{(v)}$, just as

[7] Explicitly, $(\mathcal{S}_i A_\infty)_{jk} = A^{(\infty)}_{[j-i][k-i]}$.

if they were pure states.

A.3.4 Comparison with experimental data

Figure A-2 shows a comparison between the gain measured at 30 K for the scattering-assisted design (on the left) and simulated with the periodic density matrix formalism (on the right). The band diagram is once again shown for reference. Qualitatively,

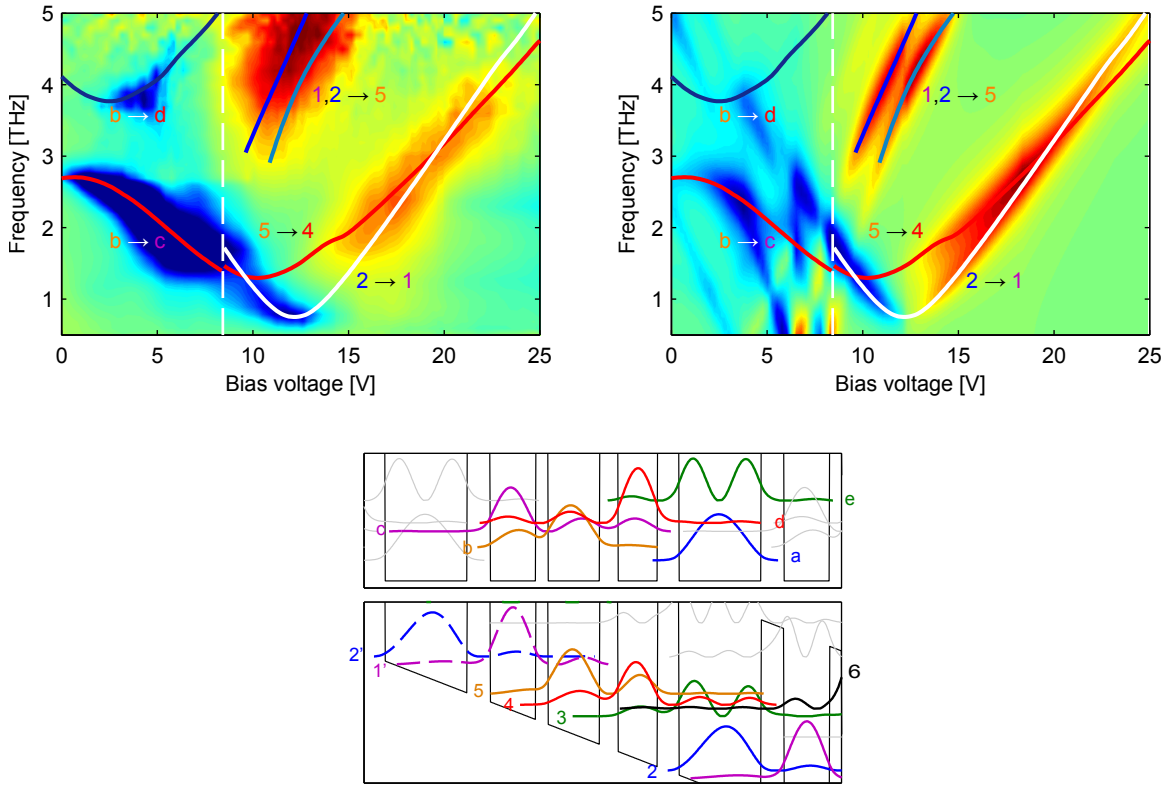


Figure A-2: Measured and simulated gain versus frequency and bias at 30 K for the OWI185E-M1 gain medium.

they agree fairly well. Below threshold the simulation predicts all sorts of absorption features that transiently appear and disappear as the bias is adjusted. Above threshold the simultation predicts both scattering-assisted gain and resonant-phonon gain. Quantitatively, there is room for improvement, e.g. the gain of the scattering-assisted transition is over-predicted. In fact, in simulations it was found to be the case that most resonant-phonon designs should have a scattering-assisted transition between

the injector and the upper laser level. This likely indicates that the optimal basis calculation did not sufficiently penalize the formation of a doublet across the injector barrier, and that perhaps higher-order calculations would be necessary to fully model this behavior.

A.3.5 Unconstrained optimization by genetic algorithms

Because the periodic density matrix formalism contains no *a priori* assumptions about the QCL band structure, it is particularly well-suited to nonlinear optimization algorithms. The key idea is to formulate a fitness function—in this case, gain at 300 K at some frequency—which is then optimized. In genetic algorithms, a population of designs is created whose fitness is evaluated at each iteration. A fittest fraction of the population is then selected for reproduction, and a new population is generated based on these reproducing designs. The new population is generated by both mutating the old designs—adding some amount of Gaussian noise—and also by mixing together layers thicknesses from different designs, much like how biological organisms combine genes. These principles were used to generate a scattering-assisted design that according to simulation should have lased up to 320 K, but in actuality only lased up to 120 K. In all likelihood, the algorithm is simply overzealous in its optimization, taking advantage of defects in the model to achieve its simulated performance.

Appendix B

Electrical modulation schemes

B.1 Asynchronous double modulation

The first method developed for determining transmission of a terahertz pulse through a QCL is referred to as *asynchronous double modulation*, and was originally developed for transmission measurements of mid-infrared QCLs [112]. In this scheme, shown in Figure B-1, the emitter section and the laser section are independently biased using

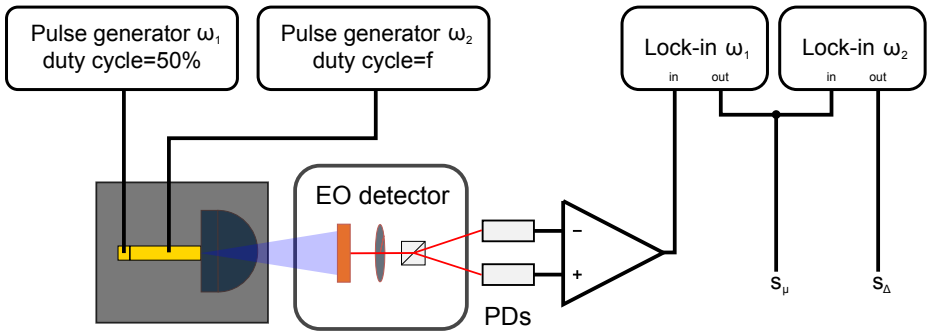


Figure B-1: Electrical schematic for asynchronous double modulation.

square waves of frequencies ω_1 and ω_2, respectively. The duty cycle of the emitter bias is usually chosen to be 50% to maximize signal, while the laser is biased at a duty cycle f. No attempt is made to synchronize their outputs. The balanced detection signal from the EO detection setup is placed into a lock-in amplifier tuned to ω_1, which simply sees the average terahertz signal produced by the emitter plus laser

system. If p_{on} and p_{off} represent the signal associated with the laser being on and the laser being off, then the output of this lock-in, denoted by s_μ, is given by

$$s_\mu = f p_{\text{on}} + (1 - f) p_{\text{off}} \tag{B.1}$$

(a weighted average). The output of this lock-in is placed into a second lock-in tuned to ω_2, which in turn represents the difference in terahertz obtained when the laser is on and when the laser is off. The resulting signal, called s_Δ, is therefore proportional to the difference signal:

$$s_\Delta = s(p_{\text{on}} - p_{\text{off}}). \tag{B.2}$$

The proportionality factor s can be determined through calibration measurements of the pulse acquired with the laser fully off. By inverting these relations, one can easily find p_{on} and p_{off} in terms of the measured parameters:

$$p_{\text{on}}(\tau) = s_\mu(\tau) + \frac{1 - f}{s}\, s_\Delta(\tau) \tag{B.3}$$

$$p_{\text{off}}(\tau) = s_\mu(\tau) - \frac{f}{s} \qquad s_\Delta(\tau). \tag{B.4}$$

One thing to note about this scheme is that since the laser usually spends much more time off than on (typical duty cycles are 10%), the SNR of $p_{\text{off}}(\tau)$ is usually much better than the SNR of $p_{\text{on}}(\tau)$. In other words, the limiting factor is the SNR of the $p_{\text{on}}(\tau)$ signal.

B.2 Synchronous double modulation

In synchronous double modulation, first performed in Ref. [68] and extended here, the laser and emitter are biased in sync with each other, with the laser being biased every other emitter pulse. The advantage of this technique is that it uses the laser duty cycle much more intelligently, because in asynchronous double modulation half the time the laser turns on the emitter is not on. In a sense, that time is completely wasted, reducing the final SNR by a factor of $\sqrt{2}$.

The procedure for performing this technique is shown in Figure B-2. Once again

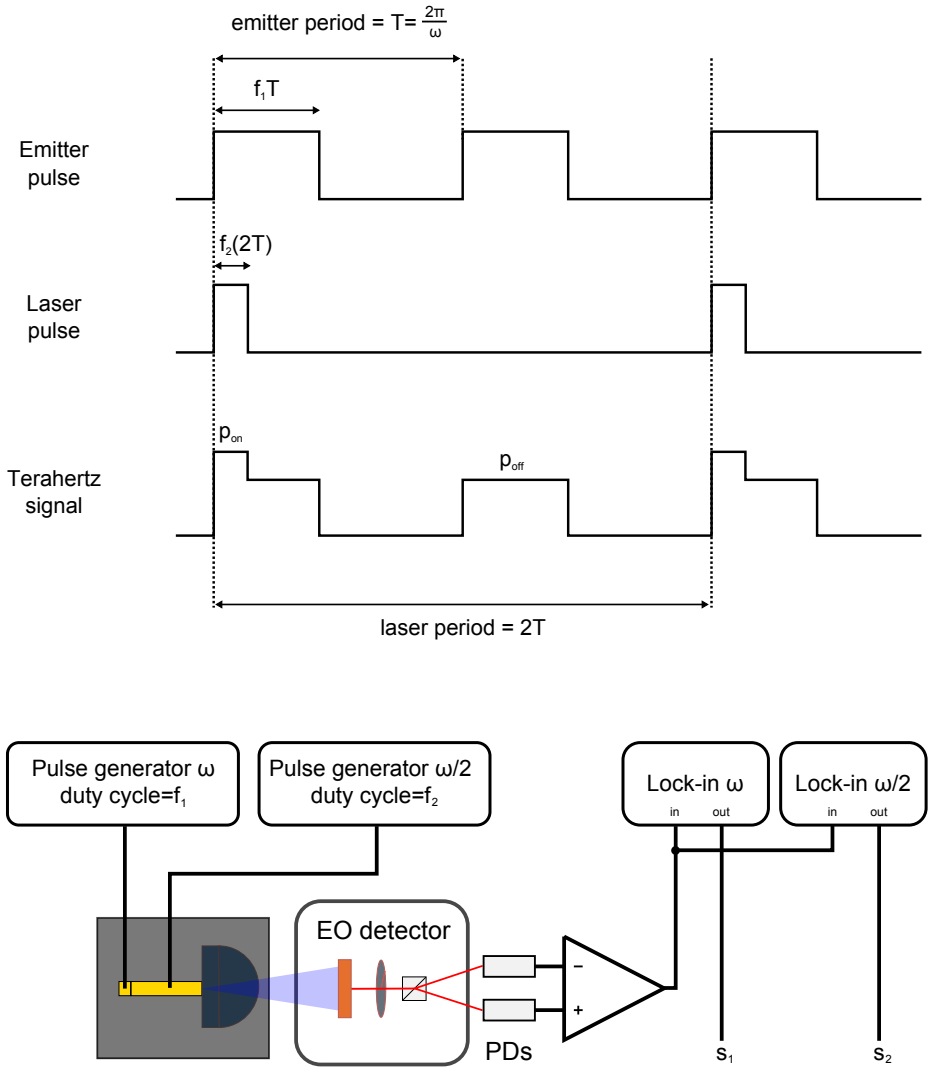


Figure B-2: Biasing scheme and electrical schematic for synchronous double modulation.

the lock-ins are locked into the emitter and laser frequencies, but this time both use the signal directly from the balanced detectors. Assume that they've been auto-phased using a TTL signal which tracks the laser and emitter bias. Let f_1 denote the duty cycle of the emitter and f_2 denote the duty cycle of the laser. One can then show using elementary Fourier analysis that the two signals detected by the lock-in

amplifiers are given by

$$\widetilde{s}_1 = p_{\text{on}} \frac{\sin(2\pi f_2)\cos((f_1 - 2f_2)\pi)}{2\sin(\pi f_1)} + p_{\text{off}} \left(1 - \frac{\sin(2\pi f_2)\cos((f_1 - 2f_2)\pi)}{2\sin(\pi f_1)}\right) \quad \text{(B.5)}$$

$$\widetilde{s}_2 = \frac{\sin(\pi f_2)}{\sin(\pi f_1)}(p_{\text{on}} - p_{\text{off}}), \quad \text{(B.6)}$$

where $\widetilde{s}_i \equiv \frac{\pi}{\sqrt{2}} \frac{1}{\sin(\pi f_1)} s_i$ is a scaled version of the raw lock-in signals. Evidently, this result is the same as the asynchronous case, with an "effective duty cycle" of

$$f = \frac{\sin(2\pi f_2)\cos((f_1 - 2f_2)\pi)}{2\sin(\pi f_1)}, \quad \text{(B.7)}$$

and an "effective scale factor" of

$$s = \frac{\sin(\pi f_2)}{\sin(\pi f_1)}. \quad \text{(B.8)}$$

Synchronous double modulation improves the SNR by a factor of $\sqrt{2}$ at the expense of some extra complexity.

Appendix C

SWIFTS analysis

C.1 Definitions and conventions

Because electric fields are continuous variables, to be fully rigorous Shifted Wave Interference Fourier Transform Spectroscopy (SWIFTS) should be derived using the language of continuous-time Fourier transforms. Conventional FTS is also analyzed here as something with which SWIFTS can be contrasted. The following conventions and definitions will be used:

1. $E(t)$ is the time-dependent electric field of the source. It will be sent through a Michelson interferometer whose variable arm is delayed by τ.

2. Even though the optical power that exits the interferometer is proportional to $P(t,\tau) = (E(t) + E(t-\tau))^2 = E^2(t) + E^2(t-\tau) + 2E(t)E(t-\tau)$, only the interference term $E(t)E(t-\tau)$ will be considered. (The first two terms are uninteresting in FTS since they are non-interferometric.)

3. Fourier transforms are referred to by capital letters, and the convention used for them is the electrical engineering one:

$$\mathscr{F}\left[f(t)\right](\omega) \equiv F(\omega) \equiv \int_{\infty}^{\infty} f(t)e^{-j\omega t}\mathrm{d}t$$

Within this convention, the well-known convolution theorem states that if $h(t) = f(t) \cdot g(t)$, then $H(\omega) = \frac{1}{2\pi} F(\omega) * G(\omega)$.

4. The effect of the final measurement of a signal $x(t)$ can be abstracted away as an integration with a measurement kernel $K(t)$. For example, if the kernel represents a simple DC average collected over a finite time interval T=1 s, then a measurement of $x(t)$ can be expressed as the integral

$$\langle x \rangle \equiv \frac{1}{T} \int_0^T x(t) \mathrm{d}t = \int_{-\infty}^{\infty} K(t) x(t) \mathrm{d}t,$$

where $K(t)$ is a boxcar function.

5. The FTS instrument used has a multplicative apodization function $A(\tau)$, a result of its finite travel range. As time-domain multiplication is frequency-domain convolution, the effect of this apodization will always be to convolve the measured interferograms with the lineshape function $A(\omega)$.

C.2 Basic analysis

C.2.1 Conventional FTS

In conventional FTS, the raw signal that is measured is essentially the field autocorrelation as a function of stage position, $S_0(\tau) = \langle E(t) E(t-\tau) \rangle$. Including the effects of apodization, its Fourier transform is given by

$$S_0(\omega) = \int_{-\infty}^{\infty} \mathrm{d}\tau A(\tau) \langle E(t) E(t-\tau) \rangle e^{-j\omega\tau}.$$

Applying the convolution theorem,

$$\begin{aligned} S_0(\omega) &= A(\omega) * \int_{-\infty}^{\infty} \mathrm{d}\tau \, \langle E(t)E(t-\tau)\rangle \, e^{-j\omega\tau} \\ &= A(\omega) * \int \mathrm{d}\tau \int \mathrm{d}t \, K(t)E(t)E(t-\tau)e^{-j\omega\tau} \\ &= A(\omega) * \int \mathrm{d}t \, K(t)E(t) \int \mathrm{d}\tau \, E(t-\tau)e^{-j\omega\tau}. \end{aligned}$$

Changing variables to $t' \equiv t - \tau$,

$$S_0(\omega) = A(\omega) * \int \mathrm{d}t \, K(t)E(t)e^{-j\omega t} \int \mathrm{d}t' \, E(t')e^{j\omega t'}.$$

The first integral is just the Fourier transform of the product of $K(t)$ and $E(t)$, and by the convolution theorem is just essentially just $(E * K)(\omega)$. The second integral is simply $E(-\omega) = E^*(\omega)$. Ignoring the factors of 2π and adding in the effect of the finite apodization, we find that the data collected by conventional FTS is

$$S_0(\omega) = A(\omega) * \Big[E^*(\omega) \cdot (E * K)(\omega)\Big] \tag{C.1}$$

Note that the measurement kernel $K(\omega)$ is typically very narrow in the frequency domain (Hz-level), and can be considered a delta function on any practical length scales. Making this substitution, one obtains the usual result that $S_0(\omega) = A(\omega) * |E(\omega)|^2$, i.e. the spectrum measured is a convolution of the power spectrum and the apodization function.

C.2.2 Shifted Wave Interference FTS

In SWIFTS, the raw signals that are collected are $S_I(\tau) = \langle E(t)E(t-\tau)\cos(\Delta\omega t)\rangle$ and $S_Q(\tau) = \langle E(t)E(t-\tau)\sin(\Delta\omega t)\rangle$ [104]. The Fourier transform of the in-phase interferogram is

$$
\begin{aligned}
S_I(\omega) &= \int_{-\infty}^{\infty} \mathrm{d}\tau A(\tau) \left\langle E(t)E(t-\tau)\cos(\Delta\omega t)\right\rangle e^{-j\omega\tau} \\
&= A(\omega) * \int \mathrm{d}\tau \int \mathrm{d}t\, K(t)E(t)E(t-\tau)\cos(\Delta\omega t)e^{-j\omega\tau} \\
&= A(\omega) * \frac{1}{2}\int \mathrm{d}t\, K(t)E(t)\left(e^{j\Delta\omega t} + e^{-j\Delta\omega t}\right)\int \mathrm{d}\tau\, E(t-\tau)e^{-j\omega\tau} \\
&= A(\omega) * \frac{1}{2}\int \mathrm{d}t\, K(t)E(t)\left(e^{-j(\omega-\Delta\omega)t} + e^{-j(\omega+\Delta\omega)t}\right)\int \mathrm{d}t'\, E(t')e^{j\omega t'} \\
&= A(\omega) * \frac{1}{2}\left[(E*K)(\omega-\Delta\omega) + (E*K)(\omega+\Delta\omega)\right]E^*(\omega).
\end{aligned}
$$

Using similar reasoning, the in-quadrature interferogram is given by

$$
S_Q(\omega) = A(\omega) * \frac{1}{2j}\left[(E*K)(\omega-\Delta\omega) - (E*K)(\omega+\Delta\omega)\right]E^*(\omega)
$$

Combining these into a complex quantity $S_\pm \equiv S_I \mp jS_Q$ and once again adding in the effect of the apodization, the final result is that

$$
S_\pm(\omega) = A(\omega) * \Big[E^*(\omega)\cdot(E*K)(\omega\pm\Delta\omega)\Big] \tag{C.2}
$$

Notice that if the measurement kernel is once again taken to be a delta function, the result simplifies to $S_\pm(\omega) = A(\omega) * \Big[E^*(\omega)E(\omega\pm\Delta\omega)\Big]$. The fact that the apodization occurs after the frequency has been shifted and multiplied is essentially the powerful aspect of SWIFTS, because it allows the user to determine how equidistant the lines of the spectrum are in a way that is independent of the instrument resolution.

C.2.3 Incoherent sources

The previous analysis implictly assumed that the electric field $E(t)$ was well-defined. For sources that are incoherent, however, one must also consider the effect of statistical fluctations on the interferograms. Conceptually, this is equivalent to considering an ensemble average of electric fields $E_i(t)$. The final results in conventional FTS and in

SWIFTS are modified by the inclusion of an average over all such ensembles:

$$S_0(\omega) = A(\omega) * \frac{1}{N}\sum_{i=1}^{N} E_i^*(\omega)E_i(\omega) \tag{C.3}$$

$$S_\pm(\omega) = A(\omega) * \frac{1}{N}\sum_{i=1}^{N} E_i^*(\omega)E_i(\omega \pm \Delta\omega). \tag{C.4}$$

For the sake of notation, the effect of the apodization and the ensemble average are frequently combined into angle brackets as follows:

$$S_0(\omega) = \Big\langle E^*(\omega)E(\omega)\Big\rangle \tag{C.5}$$

$$S_\pm(\omega) = \Big\langle E^*(\omega)E(\omega \pm \Delta\omega)\Big\rangle. \tag{C.6}$$

Note that for completely incoherent sources like glow-bars, the FTS signal will be non-zero whenever $|E(\omega)|$ is non-zero, whereas in SWIFTS the signal will vanish even when $|E(\omega)|$ and $|E(\omega \pm \Delta\omega)|$ are non-zero. This is because SWIFTS contains a phase factor, and the average over a random phase will wash out the spectrum.

C.3 Generalization to non-ideal beamsplitters

If the beamsplitter used in the Michelson interferometer is non-ideal, the preceding analysis needs to be modified. The simplest way is to allow the electric fields put into the fixed and variable arms of the interferometer to differ, calling them $E_1(t)$ and $E_2(t)$, respectively. Next, the following substitution is made:

$$E(t)E(t-\tau) \longrightarrow E_1(t)E_2(t-\tau)$$

The analysis is the same as before, with the previous results becoming

$$S_0(\omega) = A(\omega) * \Big[E_2^*(\omega) \cdot (E_1 * K)(\omega)\Big] \tag{C.7}$$

$$S_\pm(\omega) = A(\omega) * \Big[E_2^*(\omega) \cdot (E_1 * K)(\omega \pm \Delta\omega)\Big] \tag{C.8}$$

Once again, the measurement kernel is considered a delta function, in which case

$$S_0(\omega) = A(\omega) * \left[E_2^*(\omega)E_1(\omega)\right] \tag{C.9}$$

$$S_\pm(\omega) = A(\omega) * \left[E_2^*(\omega)E_1(\omega \pm \Delta\omega)\right] \tag{C.10}$$

Suppose now that $E_1(\omega) = E(\omega)$ and $E_2 = H(\omega)E(\omega)$, where H represents the transfer function of the beamsplitter. In this situation, the phase imparted by the beamsplitter causes the phase of the normal interferogram to be non-zero.

$$S_0(\omega) = A(\omega) * \left[H^*(\omega)\left|E(\omega)\right|^2\right] \tag{C.11}$$

$$S_\pm(\omega) = A(\omega) * \left[H^*(\omega)E^*(\omega)E(\omega \pm \Delta\omega)\right] \tag{C.12}$$

This is actually the reason why real interferograms are typically asymmetric. It is also the reason phase correction is often performed in commercial FTIR systems. Luckily, since the phase of the uncorrupted interferogram is known to be zero, this means that the phase of the normal interferogram can be used to correct for the phase of the SWIFT interferograms as well. This is especially critical when SWIFTS is used for phase retrieval.

C.4 Effect of demodulation imperfections

The previous analysis assumed that the in-quadrature signal was precisely given by $S_Q(\tau) = \langle E(t)E(t-\tau)\sin(\Delta\omega t)\rangle$. In general, one must actually take S_Q to be $S_Q(\tau) = \langle E(t)E(t-\tau)A_Q\cos(\Delta\omega t + \phi_Q)\rangle$, where $A_Q \approx 1$ and $\phi_Q \approx -\pi/2$. This accounts for demodulation error. In this case, one can show that the definition of $S_\pm$ needs to be modified as follows:

$$S_\pm(\omega) = \frac{S_I(\omega) - \frac{1}{A_Q}e^{\mp j\phi_Q}S_Q(\omega)}{\frac{1}{2}(1 - e^{\mp j2\phi_Q})} \tag{C.13}$$

Note that when $A_Q = 1$ and $\phi_Q = -\pi/2$, this expression reduces to the earlier definition $S_\pm = S_I \mp jS_Q$. The amplitude and phase factors can be measured directly

using calibration, or they can be measured from the non-interferometric component of SWIFTS corresponding to the variable arm.

Bibliography

[1] R. Kohler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davies, D. A. Ritchie, R. C. Iotti, and F. Rossi, "Terahertz semiconductor-heterostructure laser," *Nature*, vol. 417, pp. 156–159, May 2002.

[2] C. Meliani, G. Post, G. Rondeau, J. Decobert, W. Mouzannar, E. Dutisseuil, and R. Lefevre, "DC-92 GHz ultra-broadband high gain InP HEMT amplifier with 410 GHz gain-bandwidth product," *Electronics Letters*, vol. 38, pp. 1175–1177, Sept. 2002.

[3] W. Hafez and M. Feng, "Experimental demonstration of pseudomorphic heterojunction bipolar transistors with cutoff frequencies above 600 GHz," *Applied Physics Letters*, vol. 86, pp. 152101–152101–3, Apr. 2005.

[4] J. D. Crowley, C. Hang, R. E. Dalrymple, D. R. Tringali, F. B. Fank, L. Wandinger, and H. B. Wallace, "140 GHz indium phosphide gunn diode," *Electronics Letters*, vol. 30, pp. 499–500, Mar. 1994.

[5] I. T. Sorokina and K. L. Vodopyanov, *Solid-state mid-infrared laser sources*. Springer, 2003.

[6] B. S. Williams, "Terahertz quantum-cascade lasers," *Nat Photon*, vol. 1, no. 9, pp. 517–525, 2007.

[7] D. Woolard, E. Brown, M. Pepper, and M. Kemp, "Terahertz frequency sensing and imaging: A time of reckoning future applications?," *Proceedings of the IEEE*, vol. 93, pp. 1722–1743, Oct. 2005.

[8] T.-Y. Kao, *From high power terahertz quantum cascade lasers to terahertz light amplifiers*. Thesis, Massachusetts Institute of Technology, 2014. Thesis: Ph. D., Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2014.

[9] M. C. Kemp, P. F. Taday, B. E. Cole, J. A. Cluff, A. J. Fitzgerald, and W. R. Tribe, "Security applications of terahertz technology," vol. 5070, pp. 44–52, 2003.

[10] M. Kemp, "Explosives detection by terahertz spectroscopy—a bridge too far?," *IEEE Transactions on Terahertz Science and Technology*, vol. 1, pp. 282–292, Sept. 2011.

[11] A. J. Fitzgerald, B. E. Cole, and P. F. Taday, "Nondestructive analysis of tablet coating thicknesses using terahertz pulsed imaging," *Journal of Pharmaceutical Sciences*, vol. 94, pp. 177–183, Jan. 2005.

[12] A. W. M. Lee, *Terahertz imaging and quantum cascade laser based devices*. Thesis, Massachusetts Institute of Technology, 2010. Thesis (Ph. D.)–Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2010.

[13] H.-J. Song and T. Nagatsuma, "Present and future of terahertz communications," *IEEE Transactions on Terahertz Science and Technology*, vol. 1, pp. 256–263, Sept. 2011.

[14] R. Piesiewicz, T. Kleine-Ostmann, N. Krumbholz, D. Mittleman, M. Koch, J. Schoebel, and T. Kurner, "Short-range ultra-broadband terahertz communications: Concepts and perspectives," *IEEE Antennas and Propagation Magazine*, vol. 49, pp. 24–39, Dec. 2007.

[15] T. Kleine-Ostmann and T. Nagatsuma, "A review on terahertz communications research," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 32, pp. 143–171, Feb. 2011.

[16] Y. C. Shen, T. Lo, P. Taday, B. Cole, W. Tribe, and M. Kemp, "Detection and identification of explosives using terahertz pulsed spectroscopic imaging," *Applied Physics Letters*, vol. 86, no. 24, pp. 241116–241116–3, 2005.

[17] C. A. Schmuttenmaer, "Exploring dynamics in the far-infrared with terahertz spectroscopy," *Chemical Reviews*, vol. 104, pp. 1759–1780, Apr. 2004.

[18] Y. Hu, P. Huang, L. Guo, X. Wang, and C. Zhang, "Terahertz spectroscopic investigations of explosives," *Physics Letters A*, vol. 359, pp. 728–732, Dec. 2006.

[19] H.-W. Hubers, S. G. Pavlov, H. Richter, A. D. Semenov, L. Mahler, A. Tredicucci, H. E. Beere, and D. A. Ritchie, "High-resolution gas phase spectroscopy with a distributed feedback terahertz quantum cascade laser," *Applied Physics Letters*, vol. 89, p. 061115, Aug. 2006.

[20] T. Yasui, E. Saneyoshi, and T. Araki, "Asynchronous optical sampling terahertz time-domain spectroscopy for ultrahigh spectral resolution and rapid data acquisition," *Applied Physics Letters*, vol. 87, p. 061101, Aug. 2005.

[21] S. L. Chuang, *Physics of Optoelectronic Devices*. Wiley-Interscience, Sept. 1995.

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

[23] B. S. Williams, S. Kumar, H. Callebaut, Q. Hu, and J. L. Reno, "Terahertz quantum-cascade laser at $\lambda \approx 100$ µm using metal waveguide for mode confinement," *Applied Physics Letters*, vol. 83, pp. 2124–2126, Sept. 2003.

[24] B. S. B. S. Williams, *Terahertz quantum cascade lasers*. Thesis, Massachusetts Institute of Technology, 2003. Thesis (Ph. D.)–Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2003.

[25] V. J. Goldman, D. C. Tsui, and J. E. Cunningham, "Evidence for LO-phonon-emission-assisted tunneling in double-barrier heterostructures," *Physical Review B*, vol. 36, pp. 7635–7637, Nov. 1987.

[26] G. Scalari, L. Ajili, J. Faist, H. Beere, E. Linfield, D. Ritchie, and G. Davies, "Far-infrared ($\lambda \sim$87 µm) bound-to-continuum quantum-cascade lasers operating up to 90 k," *Applied Physics Letters*, vol. 82, pp. 3165–3167, May 2003.

[27] B. S. Williams, H. Callebaut, S. Kumar, Q. Hu, and J. L. Reno, "3.4-THz quantum cascade laser based on longitudinal-optical-phonon scattering for depopulation," *Applied Physics Letters*, vol. 82, no. 7, p. 1015, 2003.

[28] S. Kumar, Q. Hu, and J. L. Reno, "186 K operation of terahertz quantum-cascade lasers based on a diagonal design," *Applied Physics Letters*, vol. 94, no. 13, p. 131105, 2009.

[29] Y. Yao, A. J. Hoffman, and C. F. Gmachl, "Mid-infrared quantum cascade lasers," *Nature Photonics*, vol. 6, pp. 432–439, July 2012.

[30] S. Fathololoumi, E. Dupont, C. Chan, Z. Wasilewski, S. Laframboise, D. Ban, A. Matyas, C. Jirauschek, Q. Hu, and H. C. Liu, "Terahertz quantum cascade lasers operating up to $\sim$200 K with optimized oscillator strength and improved injection tunneling," *Optics Express*, vol. 20, pp. 3866–3876, Feb. 2012.

[31] S. Kumar, C. W. I. Chan, Q. Hu, and J. L. Reno, "A 1.8-THz quantum cascade laser operating significantly above the temperature of $\hbar\omega/k_b$," *Nature Physics*, vol. 7, pp. 166–171, Feb. 2011.

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

[33] M. S. Vitiello, G. Scamarcio, V. Spagnolo, B. S. Williams, S. Kumar, Q. Hu, and J. L. Reno, "Measurement of subband electronic temperatures and population inversion in THz quantum-cascade lasers," *Applied Physics Letters*, vol. 86, no. 11, p. 111115, 2005.

[34] S. Kumar and Q. Hu, "Coherence of resonant-tunneling transport in terahertz quantum-cascade lasers," *Physical Review B*, vol. 80, p. 245316, Dec. 2009.

[35] R. Nelander, A. Wacker, M. F. Pereira, D. G. Revin, M. R. Soulby, L. R. Wilson, J. W. Cockburn, A. B. Krysa, J. S. Roberts, and R. J. Airey, "Fingerprints of spatial charge transfer in quantum cascade lasers," *Journal of Applied Physics*, vol. 102, pp. 113104–113104–5, Dec. 2007.

[36] D. O. Winge, M. Lindskog, and A. Wacker, "Nonlinear response of quantum cascade structures," *Applied Physics Letters*, vol. 101, p. 211113, Nov. 2012.

[37] A. Wacker, G. Bastard, F. Carosella, R. Ferreira, and E. Dupont, "Unraveling of free-carrier absorption for terahertz radiation in heterostructures," *Physical Review B*, vol. 84, p. 205319, Nov. 2011.

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

[39] J. Kroll, J. Darmo, S. S. Dhillon, X. Marcadet, M. Calligaro, C. Sirtori, and K. Unterrainer, "Phase-resolved measurements of stimulated emission in a laser," *Nature*, vol. 449, pp. 698–701, Oct. 2007.

[40] J. Kroll, J. Darmo, K. Unterrainer, S. S. Dhillon, C. Sirtori, X. Marcadet, and M. Calligaro, "Longitudinal spatial hole burning in terahertz quantum cascade lasers," *Applied Physics Letters*, vol. 91, pp. 161108–3, Oct. 2007.

[41] S. Kohen, B. S. Williams, and Q. Hu, "Electromagnetic modeling of terahertz quantum cascade laser waveguides and resonators," *Journal of Applied Physics*, vol. 97, no. 5, p. 053106, 2005.

[42] P. Smith, D. Auston, and M. Nuss, "Subpicosecond photoconducting dipole antennas," *IEEE Journal of Quantum Electronics*, vol. 24, pp. 255–260, Feb. 1988.

[43] A. Nahata, A. S. Weling, and T. F. Heinz, "A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling," *Applied Physics Letters*, vol. 69, no. 16, p. 2321, 1996.

[44] N. Karpowicz, J. Dai, X. Lu, Y. Chen, M. Yamaguchi, H. Zhao, X.-C. Zhang, L. Zhang, C. Zhang, M. Price-Gallagher, C. Fletcher, O. Mamer, A. Lesimple, and K. Johnson, "Coherent heterodyne time-domain spectrometry covering the entire "terahertz gap"," *Applied Physics Letters*, vol. 92, p. 011131, Jan. 2008.

[45] K.-L. Yeh, M. C. Hoffmann, J. Hebling, and K. A. Nelson, "Generation of 10 µJ ultrashort terahertz pulses by optical rectification," *Applied Physics Letters*, vol. 90, p. 171121, Apr. 2007.

[46] P. Gu, M. Tani, S. Kono, K. Sakai, and X.-C. Zhang, "Study of terahertz radiation from InAs and InSb," *Journal of Applied Physics*, vol. 91, pp. 5533–5537, May 2002.

[47] G. Klatt, F. Hilser, W. Qiao, M. Beck, R. Gebs, A. Bartels, K. Huska, U. Lemmer, G. Bastian, M. Johnston, M. Fischer, J. Faist, and T. Dekorsy, "Terahertz emission from lateral photo-dember currents," *Optics Express*, vol. 18, p. 4939, Mar. 2010.

[48] S. Gupta, M. Y. Frankel, J. A. Valdmanis, J. F. Whitaker, G. A. Mourou, F. W. Smith, and A. R. Calawa, "Subpicosecond carrier lifetime in GaAs grown by molecular beam epitaxy at low temperatures," *Applied Physics Letters*, vol. 59, pp. 3276–3278, Dec. 1991.

[49] G. Rebeiz, "Millimeter-wave and terahertz integrated circuit antennas," *Proceedings of the IEEE*, vol. 80, pp. 1748–1770, Nov. 1992.

[50] K. J. Button, *Infrared and Millimeter Waves V10: Millimeter Components and Techniques*. Elsevier, Jan. 1983.

[51] A. Wei Min Lee, Q. Qin, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno, "High-power and high-temperature THz quantum-cascade lasers based on lens-coupled metal-metal waveguides," *Optics Letters*, vol. 32, pp. 2840–2842, Oct. 2007.

[52] Y. C. Shen, P. C. Upadhya, E. H. Linfield, H. E. Beere, and A. G. Davies, "Ultrabroadband terahertz radiation from low-temperature-grown GaAs photoconductive emitters," *Applied Physics Letters*, vol. 83, no. 15, p. 3117, 2003.

[53] L. Liu, J. Hesler, H. Xu, A. Lichtenberger, and R. Weikle, "A broadband quasi-optical terahertz detector utilizing a zero bias schottky diode," *IEEE Microwave and Wireless Components Letters*, vol. 20, pp. 504–506, Sept. 2010.

[54] W. J. Moore and H. Shenker, "A high-detectivity gallium-doped germanium detector for the 40-120μ region," *Infrared Physics*, vol. 5, pp. 99–106, Sept. 1965.

[55] W. Zhang, P. Khosropanah, J. R. Gao, E. L. Kollberg, K. S. Yngvesson, T. Bansal, R. Barends, and T. M. Klapwijk, "Quantum noise in a terahertz hot electron bolometer mixer," *Applied Physics Letters*, vol. 96, p. 111113, Mar. 2010.

[56] Q. Wu and X.-C. Zhang, "Free-space electro-optic sampling of terahertz beams," *Applied Physics Letters*, vol. 67, pp. 3523–3525, Dec. 1995.

[57] Q. Wu and X.-C. Zhang, "7 terahertz broadband GaP electro-optic sensor," *Applied Physics Letters*, vol. 70, pp. 1784–1786, Apr. 1997.

[58] Q. Wu, M. Litz, and X.-C. Zhang, "Broadband detection capability of ZnTe electro-optic field detectors," *Applied Physics Letters*, vol. 68, no. 21, p. 2924, 1996.

[59] Q. Wu and X.-C. Zhang, "Free-space electro-optics sampling of mid-infrared pulses," *Applied Physics Letters*, vol. 71, no. 10, pp. 1285–1286, 1997.

[60] M. Xiao, L.-A. Wu, and H. J. Kimble, "Precision measurement beyond the shot-noise limit," *Physical Review Letters*, vol. 59, pp. 278–281, July 1987.

[61] N. Jukam, S. S. Dhillon, D. Oustinov, Z.-Y. Zhao, S. Hameau, J. Tignon, S. Barbieri, A. Vasanelli, P. Filloux, C. Sirtori, and X. Marcadet, "Investigation of spectral gain narrowing in quantum cascade lasers using terahertz time domain spectroscopy," *Applied Physics Letters*, vol. 93, no. 10, pp. 101115–3, 2008.

[62] J. Lloyd-Hughes, G. Scalari, A. van Kolck, M. Fischer, M. Beck, and J. Faist, "Coupling terahertz radiation between sub-wavelength metal-metal waveguides and free space using monolithically integrated horn antennae," *Optics Express*, vol. 17, no. 20, pp. 18387–18393, 2009.

[63] S. S. Dhillon, S. Sawallich, N. Jukam, D. Oustinov, J. Madeo, S. Barbieri, P. Filloux, C. Sirtori, X. Marcadet, and J. Tignon, "Integrated terahertz pulse generation and amplification in quantum cascade lasers," *Applied Physics Letters*, vol. 96, no. 6, p. 061107, 2010.

[64] M. Martl, J. Darmo, D. Dietze, K. Unterrainer, and E. Gornik, "Terahertz waveguide emitter with subwavelength confinement," *Journal of Applied Physics*, vol. 107, no. 1, p. 013110, 2010.

[65] N. Jukam, S. S. Dhillon, D. Oustinov, J. Madeo, J. Tignon, R. Colombelli, P. Dean, M. Salih, S. P. Khanna, E. H. Linfield, and A. G. Davies, "Terahertz time domain spectroscopy of phonon-depopulation based quantum cascade lasers," *Applied Physics Letters*, vol. 94, no. 25, p. 251108, 2009.

[66] D. Burghoff, T.-Y. Kao, D. Ban, A. W. M. Lee, Q. Hu, and J. Reno, "A terahertz pulse emitter monolithically integrated with a quantum cascade laser," *Applied Physics Letters*, vol. 98, no. 6, p. 061112, 2011.

[67] D. Burghoff, C. Wang Ivan Chan, Q. Hu, and J. L. Reno, "Gain measurements of scattering-assisted terahertz quantum cascade lasers," *Applied Physics Letters*, vol. 100, pp. 261111–261111–4, June 2012.

[68] N. Jukam, S. Dhillon, Z.-Y. Zhao, G. Duerr, J. Armijo, N. Sirmons, S. Hameau, S. Barbieri, P. Filloux, C. Sirtori, X. Marcadet, and J. Tignon, "Gain measurements of THz quantum cascade lasers using THz time-domain spectroscopy," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 14, no. 2, pp. 436–442, 2008.

[69] D. Oustinov, N. Jukam, R. Rungsawang, J. Madéo, S. Barbieri, P. Filloux, C. Sirtori, X. Marcadet, J. Tignon, and S. Dhillon, "Phase seeding of a terahertz quantum cascade laser," *Nature Communications*, vol. 1, no. 6, p. 69, 2010.

[70] B. Williams, S. Kumar, Q. Hu, and J. Reno, "Resonant-phonon terahertz quantum-cascade laser operating at 2.1 THz ($\lambda \approx$141μm)," *Electronics Letters*, vol. 40, no. 7, pp. 431–433, 2004.

[71] S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno, "1.9 THz quantum-cascade lasers with one-well injector," *Applied Physics Letters*, vol. 88, no. 12, p. 121123, 2006.

[72] M. Yamanishi, K. Fujita, N. Yu, T. Edamura, K. Tanaka, and F. Capasso, "3-4 THz InGaAs/InAlAs quantum-cascade lasers based on the indirect pump scheme," in *CLEO: Science and Innovations*, 2011.

[73] E. Dupont, S. Fathololoumi, Z. R. Wasilewski, G. Aers, S. R. Laframboise, M. Lindskog, S. G. Razavipour, A. Wacker, D. Ban, and H. C. Liu, "A phonon scattering assisted injection and extraction based terahertz quantum cascade laser," *Journal of Applied Physics*, vol. 111, pp. 073111–073111–10, Apr. 2012.

[74] D. Turčinková, G. Scalari, F. Castellano, M. I. Amanti, M. Beck, and J. Faist, "Ultra-broadband heterogeneous quantum cascade laser emitting from 2.2 to 3.2 THz," *Applied Physics Letters*, vol. 99, pp. 191104–191104–3, Nov. 2011.

[75] M. Rösch, G. Scalari, M. Beck, and J. Faist, "Octave-spanning semiconductor laser," *arXiv preprint arXiv:1407.3050*, 2014.

[76] J. R. Freeman, A. Brewer, J. Madéo, P. Cavalié, S. S. Dhillon, J. Tignon, H. E. Beere, and D. A. Ritchie, "Broad gain in a bound-to-continuum quantum cascade laser with heterogeneous active region," *Applied Physics Letters*, vol. 99, pp. 241108–241108–3, Dec. 2011.

[77] A. Wei Min Lee, T.-Y. Kao, D. Burghoff, Q. Hu, and J. L. Reno, "Terahertz tomography using quantum-cascade lasers," *Optics Letters*, vol. 37, pp. 217–219, Jan. 2012.

[78] M. Kourogi, K. Nakagawa, and M. Ohtsu, "Wide-span optical frequency comb generator for accurate optical frequency difference measurement," *IEEE Journal of Quantum Electronics*, vol. 29, no. 10, pp. 2693–2701, 1993.

[79] J. C. Pearson, B. J. Drouin, A. Maestrini, I. Mehdi, J. Ward, R. H. Lin, S. Yu, J. J. Gill, B. Thomas, C. Lee, G. Chattopadhyay, E. Schlecht, F. W. Maiwald, P. F. Goldsmith, and P. Siegel, "Demonstration of a room temperature 2.48-2.75 THz coherent spectroscopy source," *Review of Scientific Instruments*, vol. 82, p. 093105, Sept. 2011.

[80] B. Drouin, F. W. Maiwald, and J. Pearson, "Application of cascaded frequency multiplication to molecular spectroscopy," *Review of Scientific Instruments*, vol. 76, no. 9, pp. 093113–093113–10, 2005.

[81] E. P. Ippen, "Principles of passive mode locking," *Applied Physics B*, vol. 58, pp. 159–170, Mar. 1994.

[82] C. Y. Wang, L. Diehl, A. Gordon, C. Jirauschek, F. X. Kärtner, A. Belyanin, D. Bour, S. Corzine, G. Höfler, M. Troccoli, J. Faist, and F. Capasso, "Coherent instabilities in a semiconductor laser with fast gain recovery," *Physical Review A*, vol. 75, p. 031802, Mar. 2007.

[83] C. Y. Wang, L. Kuznetsova, V. M. Gkortsas, L. Diehl, F. X. Kärtner, M. A. Belkin, A. Belyanin, X. Li, D. Ham, H. Schneider, P. Grant, C. Y. Song, S. Haffouz, Z. R. Wasilewski, H. Liu, and F. Capasso, "Mode-locked pulses from mid-infrared quantum cascade lasers," *Optics Express*, vol. 17, pp. 12929–12943, July 2009.

[84] S. Barbieri, M. Ravaro, P. Gellie, G. Santarelli, C. Manquest, C. Sirtori, S. P. Khanna, E. H. Linfield, and A. G. Davies, "Coherent sampling of active mode-locked terahertz quantum cascade lasers and frequency synthesis," *Nature Photonics*, vol. 5, no. 5, pp. 306–313, 2011.

[85] A. Hugi, G. Villares, S. Blaser, H. C. Liu, and J. Faist, "Mid-infrared frequency comb based on a quantum cascade laser," *Nature*, vol. 492, pp. 229–233, Dec. 2012.

[86] T. J. Kippenberg, R. Holzwarth, and S. A. Diddams, "Microresonator-based optical frequency combs," *Science*, vol. 332, pp. 555–559, Apr. 2011.

[87] F. Keilmann, C. Gohle, and R. Holzwarth, "Time-domain mid-infrared frequency-comb spectrometer," *Optics Letters*, vol. 29, pp. 1542–1544, July 2004.

[88] B. Bernhardt, A. Ozawa, P. Jacquet, M. Jacquey, Y. Kobayashi, T. Udem, R. Holzwarth, G. Guelachvili, T. W. Hänsch, and N. Picqué, "Cavity-enhanced dual-comb spectroscopy," *Nature Photonics*, vol. 4, pp. 55–57, Jan. 2010.

[89] P. Del'Haye, A. Schliesser, O. Arcizet, T. Wilken, R. Holzwarth, and T. J. Kippenberg, "Optical frequency comb generation from a monolithic microresonator," *Nature*, vol. 450, pp. 1214–1217, Dec. 2007.

[90] J. S. Levy, A. Gondarenko, M. A. Foster, A. C. Turner-Foster, A. L. Gaeta, and M. Lipson, "CMOS-compatible multiple-wavelength oscillator for on-chip optical interconnects," *Nature Photonics*, vol. 4, pp. 37–40, Jan. 2010.

[91] L. Razzari, D. Duchesne, M. Ferrera, R. Morandotti, S. Chu, B. E. Little, and D. J. Moss, "CMOS-compatible integrated optical hyper-parametric oscillator," *Nature Photonics*, vol. 4, pp. 41–45, Jan. 2010.

[92] R. W. Boyd, *Nonlinear Optics, Third Edition*. Amsterdam; Boston: Academic Press, third ed., Apr. 2008.

[93] A. E. Siegman, *Lasers*. Mill Valley, Calif.: University Science Books, May 1986.

[94] M. S. Dresselhaus, "Optical properties of solids,"

[95] J. S. Blakemore, "Semiconducting and other major properties of gallium arsenide," *Journal of Applied Physics*, vol. 53, pp. R123–R181, Oct. 1982.

[96] F. X. Kaertner, U. Morgner, R. Ell, T. Schibli, J. G. Fujimoto, E. P. Ippen, V. Scheuer, G. Angelow, and T. Tschudi, "Ultrabroadband double-chirped mirror pairs for generation of octave spectra," *Journal of the Optical Society of America B*, vol. 18, pp. 882–885, June 2001.

[97] P. Gellie, S. Barbieri, J.-F. Lampin, P. Filloux, C. Manquest, C. Sirtori, I. Sagnes, S. P. Khanna, E. H. Linfield, A. G. Davies, H. Beere, and D. Ritchie, "Injection-locking of terahertz quantum cascade lasers up to 35ghz using RF amplitude modulation," *Optics Express*, vol. 18, no. 20, pp. 20799–20816, 2010.

[98] V. Torres-Company, D. Castello-Lurbe, and E. Silvestre, "Comparative analysis of spectral coherence in microresonator frequency combs," *Optics Express*, vol. 22, p. 4678, Feb. 2014.

[99] R. P. Green, A. Tredicucci, N. Q. Vinh, B. Murdin, C. Pidgeon, H. E. Beere, and D. A. Ritchie, "Gain recovery dynamics of a terahertz quantum cascade laser," *Physical Review B*, vol. 80, no. 7, p. 075303, 2009.

[100] M. C. Wanke, E. W. Young, C. D. Nordquist, M. J. Cich, A. D. Grine, C. T. Fuller, J. L. Reno, and M. Lee, "Monolithically integrated solid-state terahertz transceivers," *Nat Photon*, vol. 4, no. 8, pp. 565–569, 2010.

[101] E. Rosencher and P. Bois, "Model system for optical nonlinearities: Asymmetric quantum wells," *Physical Review B*, vol. 44, pp. 11315–11327, Nov. 1991.

[102] A. Baryshev, J. N. Hovenier, A. J. L. Adam, I. Kašalynas, J. R. Gao, T. O. Klaassen, B. S. Williams, S. Kumar, Q. Hu, and J. L. Reno, "Phase locking and spectral linewidth of a two-mode terahertz quantum cascade laser," *Applied Physics Letters*, vol. 89, p. 031115, July 2006.

[103] A. Gordon, C. Y. Wang, L. Diehl, F. X. Kärtner, A. Belyanin, D. Bour, S. Corzine, G. Höfler, H. C. Liu, H. Schneider, T. Maier, M. Troccoli, J. Faist, and F. Capasso, "Multimode regimes in quantum cascade lasers: From coherent instabilities to spatial hole burning," *Physical Review A*, vol. 77, p. 053804, May 2008.

[104] D. Burghoff, T.-Y. Kao, N. Han, C. W. I. Chan, X. Cai, Y. Yang, D. J. Hayton, J.-R. Gao, J. L. Reno, and Q. Hu, "Terahertz laser frequency combs," *Nature Photonics*, vol. 8, pp. 462–467, June 2014.

[105] R. Trebino, K. W. DeLong, D. N. Fittinghoff, J. N. Sweetser, M. A. KrumbÃijgel, B. A. Richman, and D. J. Kane, "Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating," *Review of Scientific Instruments*, vol. 68, pp. 3277–3295, Sept. 1997.

[106] C. Iaconis and I. Walmsley, "Spectral phase interferometry for direct electric-field reconstruction of ultrashort optical pulses," *Optics Letters*, vol. 23, pp. 792–794, May 1998.

[107] J. B. Khurgin, Y. Dikmelik, A. Hugi, and J. Faist, "Coherent frequency combs produced by self frequency modulation in quantum cascade lasers," *Applied Physics Letters*, vol. 104, p. 081118, Feb. 2014.

[108] T.-Y. Kao, Q. Hu, and J. L. Reno, "Perfectly phase-matched third-order distributed feedback terahertz quantum-cascade lasers," *Optics Letters*, vol. 37, pp. 2070–2072, June 2012.

[109] M. I. Amanti, M. Fischer, G. Scalari, M. Beck, and J. Faist, "Low-divergence single-mode terahertz quantum cascade laser," *Nature Photonics*, vol. 3, no. 10, pp. 586–590, 2009.

[110] D. J. Hayton, A. Khudchenko, D. G. Pavelyev, J. N. Hovenier, A. Baryshev, J. R. Gao, T. Y. Kao, Q. Hu, J. L. Reno, and V. Vaks, "Phase locking of a 3.4 THz third-order distributed feedback quantum cascade laser using a room-temperature superlattice harmonic mixer," *Applied Physics Letters*, vol. 103, p. 051115, Aug. 2013.

[111] Y. Ren, J. N. Hovenier, M. Cui, D. J. Hayton, J. R. Gao, T. M. Klapwijk, S. C. Shi, T.-Y. Kao, Q. Hu, and J. L. Reno, "Frequency locking of single-mode 3.5-THz quantum cascade lasers using a gas cell," *Applied Physics Letters*, vol. 100, p. 041111, Jan. 2012.

[112] D. Burghoff, *Characterization of mid-infrared quantum cascade lasers*. Thesis, Massachusetts Institute of Technology, 2009. Thesis (S.M.)–Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2009.

[113] T. E. Abrudan, J. Eriksson, and V. Koivunen, "Steepest descent algorithms for optimization under unitary matrix constraint," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1134–1147, Mar. 2008.