Characterization and Analysis of Highly Diagonal Terahertz Quantum Cascade Lasers

by

Chun Wang Ivan Chan

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

Master of Science in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author .....
Department of Electrical Engineering and Computer Science
August 25, 2010

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

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

Characterization and Analysis of Highly Diagonal Terahertz Quantum Cascade Lasers

by

Chun Wang Ivan Chan

Submitted to the Department of Electrical Engineering and Computer Science
on August 25, 2010, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering

Abstract

The as yet unattained milestone of room-temperature operation is essential for establishing Terahertz Quantum Cascade Lasers (THz QCLs) as practical sources of THz radiation. Temperature performance is hypothesized to be limited by upper laser level lifetime reduction due to non-radiative scattering, particularly by longitudinal optical phonons. To address this issue, this work studies highly "diagonal" QCLs, where the upper and lower laser level wavefunctions are spatially separated to preserve upper laser level lifetime, as well as several other issues relevant to high temperature performance.

The highly diagonal devices of this work performed poorly, but the analysis herein nevertheless suggest that diagonality as a design strategy cannot yet be ruled out. Other causes of poor performance in the lasers are identified, and suggestions for future designs are made.

Thesis Supervisor: Qing Hu
Title: Professor

Acknowledgments

Many are those whom I must thank. I thank my supervisor Professor Qing Hu for his guidance and support over the last two years. I thank Dr. Sushil Kumar for this invaluable tutelage; I aspire to his patience, should the day come when I am called to teach as well. I also thank my lab mates Dr. Alan Lee, Qi Qin, Wilt Kao, and David Burghoff for being an excellent gang of friends, ever ready to share in jokes and ideas. I am both humbled and honored to share in the company of so much good will and talent. Additionally, I thank Dr. John Reno at Sandia National Laboratories for the MBE growth crucial to this work.

Outside of academia, I am indebted to my family for a quarter-century of unwavering love and support. In particular, if I have succeeded at all in life, I owe everything to my grandmother, Sik-Je Lam, who unfortunately passed away February of this year; I will never forget her.

Finally, I gratefully acknowledge funding from NASA, the primary sponsor of this research, and the National Science and Engineering Research Council of Canada, who provided me with my first year scholarship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The two major sources of coherent radiation in modern technologies are electronic oscillators and conventional lasers. The need for explicit charge transfer limits electronic oscillators via RC time constants to approximately <300GHz. Conversely, naturally occurring bandgaps tend not to fall below ∼60meV (eg. in lead-salt lasers [1]), limiting conventional lasers to approximately >15THz. The source-poor region of the spectrum lying in ∼0.3THz–10THz is therefore known as the *terahertz gap* (see figure 1-1). Diverse applications ranging from chemical sensing and spectroscopy to security and astronomical imaging have been proposed for the THz spectrum (see, for example, [2–6]), thus it remains desirable to find compact sources of continuous-wave, high power, coherent THz radiation.

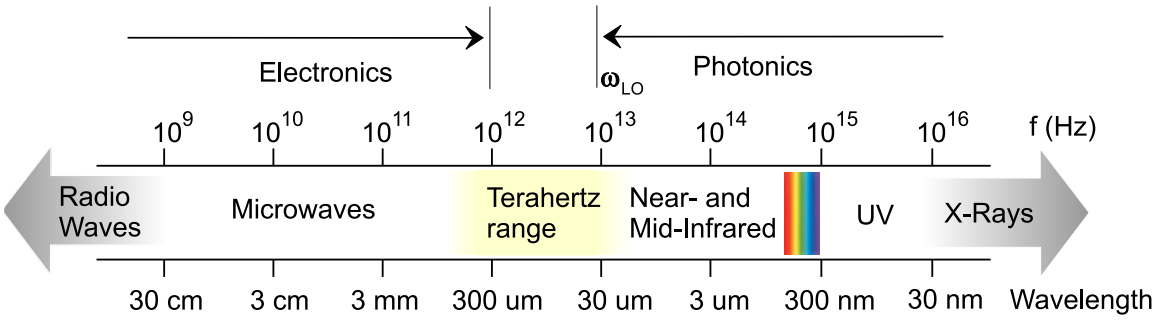


Figure 1-1: The "terahertz gap" in the electromagnetic spectrum. Few natural sources of radiation exist in this range.

At present, no entirely satisfactory source exists. Electronic sources such as Gunn

oscillators and Schottky-diode frequency multipliers provide output powers only of order $\sim 10\mu$W [7, 8]. Solid-state lasers based on strained p-type germanium with crossed electric and magnetic fields or optically pumped impurity state transitions in silicon require liquid helium operating temperatures and permit only pulsed operation due to large energy consumption and poor efficiency [9]. Optically pumped molecular gas lasers based on vibrational/rotational molecular state transitions are bulky and energy inefficient, and moreover yield only a limited number of frequencies (although they have seen some practical use [10]). Free electron lasers offer high power and broad tuning, but are even more unwieldy; they remain impractical outside of research due to large cost, size, and infrastructure requirements.

The topic of this thesis is the *Quantum Cascade Laser* (QCL), arguably the most promising THz source currently under development.

1.1 Quantum Cascade Lasers

At heart, a QCL is a *superlattice*. First proposed by Esaki and Tsu in 1970 [11], a superlattice is a periodic stack of semiconductor films of varying thicknesses that is typically grown by molecular beam epitaxy (MBE). At the heterojunctions boundaries between different semiconductor layers, the abrupt change in lattice potential creates discontinuities in the conduction band-edge energies, leading to quantum confinement in the growth direction. This splits the material conduction band states into *subbands*, between which optical transitions can occur (see figure 1-2).

Kazarinov and Suris first proposed the basic idea of a QCL in 1971 [12]: through careful selection of layer widths, a biased superlattice can achieve optical gain through population inversion between subband states. Referring to figure 1-3, a single electron can "cascade" in energy down the superlattice while emitting a photon in each superlattice period, ingeniously enabling effective quantum efficiencies much higher than unity (in principle).

The first QCL was demonstrated in 1994, lasing in the mid-infrared (mid-IR) [13]. Mid-IR QCLs have since advanced rapidly in power, efficiency, and maximum lasing

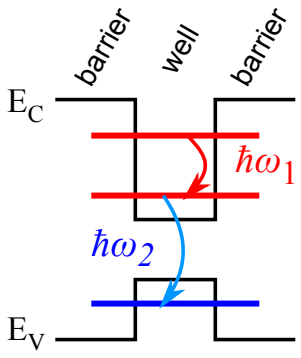


Figure 1-2: Electronic transitions in a heterostructure. The abrupt material discontinuity leads to both conduction and valence band energy level discretizations in the well region. $\hbar\omega_1$ is an *intersubband* transition, on which QCLs are based, and $\hbar\omega_2$ is an *interband* transition, on which traditional semiconductor lasers are based.

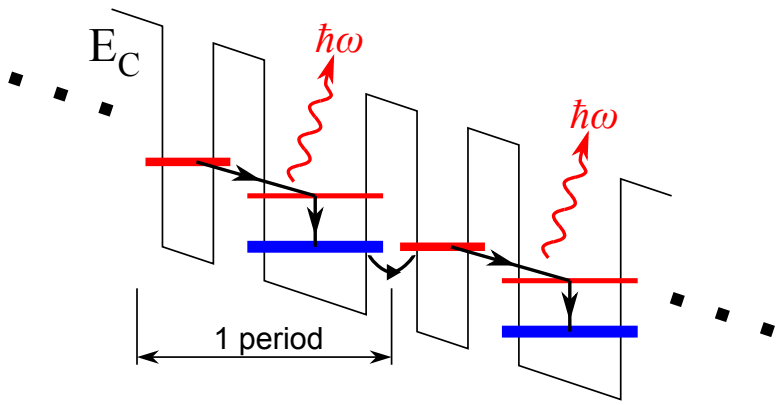


Figure 1-3: Schematic of QCL operation. In principle, each electron emits one photon in each superlattice period before moving into the next period.

temperature [14], establishing themselves in diverse applications such as spectroscopy and optical communications [15]. Köhler et al. demonstrated the first operational THz QCL in 2002 [16]. THz QCLs now cover the spectral range from 5THz down to 1.2THz (or even lower with magnetic field assistance [17, 18]), and can emit up to hundreds of milliwatts of power.

1.2 THz QCL Temperature Performance

Unfortunately, THz QCLs have in general lagged behind their mid-IR counterparts in progress; perhaps the eight year gap between the first mid-IR QCL and the first THz QCL attests to the difficulty of working in the THz regime. The foremost barrier to application is that THz QCLs require cryogenic operation. Whereas mid-IR QCLs have long ago attained continuous-wave room temperature operation [19], no THz QCL to date operates above 186K (see design OWI222G in section 4.2.5, and also [20]). This thesis details theoretical and experimental investigations into the limits to temperature performance and progress made towards achieving higher temperature operation.

The postulated limit to high temperature performance is degradation of population inversion due to non-radiative scattering. With increasing temperature, scattering mechanisms (most notably interactions with LO phonons) cause upper laser subband electrons to transit to the lower laser subband without photon emission. Section 3.2.1 later in this thesis elaborates further on this topic.

Solutions such as magnetic confinement [18, 21], zero-dimensional heterostructures [22], and new material systems (in particular, GaAsN [23]) have been proposed in the literature to quench non-radiative scattering mechanisms. In particular, magnetic confinement has already proven capable of raising operating temperatures to 225K, but the equipment needed to generate the requisite ∼10-16T magnetic fields is even less practical than cryogenic cooling equipment. Adequate zero-dimensional heterostructures have yet to be fabricated, and epitaxial growth in GaAsN is notoriously difficult [24].

As an alternative, the MIT THz QCL group of Professor Qing Hu is pursuing the less radical strategy of increasing the spatial separation between upper and lower laser level wavefunctions, or the so-called " diagonality" of the laser design. While high diagonality carries its own disadvantages, calculated scattering rates indicate that diagonality ensures the survival of population inversion up to room temperatures. The thesis explores the effects of diagonality on THz QCL design, particularly on observed experimental transport properties.

1.3 Thesis Overview

The remainder of this thesis is organized as follows.

- Chapter 2 expands upon the qualitative introduction given in this chapter, explaining the physical models of QCL transport.
- Chapter 3 draws upon the theoretical foundation formed in chapter 2 to discuss QCL design.
- Chapter 4 presents the analysis and experimental data for the devices studied, and draws final conclusions.

Chapter 2

The Theory of Quantum Cascade Laser Operation

Although conceptually easy to understand, QCL operation is challenging to model quantitatively. This is because electrical transport in a QCL is neither fully ballistic nor fully scattering-assisted, so traditional approaches such as Landauer-Buttiker transmission and the drift-diffusion equations do not apply. This chapter details the semiclassical/quantum phenomenological methods that have been adapted for QCL analysis. One must bear in mind that these methods, although useful, are highly approximate.

In recent years, much effort has been devoted to the development of fully quantum mechanical theories of transport in the formalism of *Nonequilibrium Green's Functions* (NEGF; see for example [25, 26]). This field of research is promising, but the tremendous computational burden of NEGF prevents its regular use in design, and even the best implementations presently available do not make consistently accurate predictions.

There are two crucial transport processes which must be considered in detail: carrier scattering and resonant tunneling.

2.1 The Envelope Function Approximation (EFA) for Superlattices

Quantum cascade lasers are typically analyzed using effective-mass theory in the envelope function approximation. There is extant work on QCL modeling using more accurate microscopic approaches, such as the tight-binding approximation [27], but the envelope function approximation remains invaluable for its simplicity and ability to give rapid results.

The governing equation is the *effective mass envelope function equation* that determines the wavefunction envelope $F(\overline{r})$,

$$-\nabla\frac{\hbar^2}{2m^*(\overline{r})}\nabla F(\overline{r}) + V(\overline{r})F(\overline{r}) = EF(\overline{r}) \tag{2.1}$$

where $m^*(\overline{r})$ is a spatially varying effective mass encoding the effects of the semiconductor crystal lattice on electron motion, and $V(\overline{r})$ is the combination of potentials due to conduction band-offsets of the superlattice and any externally applied electrical potential. More generally, $1/m^*$ can be a matrix for anisotropic bands, but this thesis considers only the conduction band of $Al_xGa_{1-x}As$, which is isotropic. Except for the spatially varying effective mass, one notes that equation (2.1) is formally identical to Schrödinger's equation, so most quantum mechanical techniques apply directly. Nevertheless, one should remember that $F(\overline{r})$ is *not* the wavefunction per se; that is given by $\Psi(\overline{r}) \approx F(\overline{r})u_{\overline{k}=0}(\overline{r})$ where $u_{\overline{k}=0}(\overline{r})$ is the zone-center Bloch amplitude.

Let the epitaxial growth direction of the QCL superlattice be denoted as $\hat{z}$. For a QCL, one typically assumes that $x-y$ plane is homogeneous and practically infinite. This is justified by the large in-plane dimensions of real devices compared to the quantum well dimensions along $\hat{z}$. Therefore, $m^*(\overline{r}) = m^*(z)$ and $V(\overline{r}) = V(z)$, and one may assume plane wave solutions in-plane, namely

$$F(\overline{r}) = F(z)\frac{e^{i\overline{k}\cdot\overline{\rho}}}{\sqrt{A}} \tag{2.2}$$

where $\overline{\rho} = x\hat{x} + y\hat{y}$, and $\overline{k}$ is the *in-plane* (not bulk) wavevector. Substituting this into equation (2.1) yields

$$\left[-\frac{\hbar^2}{2}\frac{d}{dz}\frac{1}{m^*(z)}\frac{d}{dz} + V(z) + \frac{\hbar^2 k^2}{2m^*(z)}\right] F_n(z) = E_n\left(\overline{k}\right) F_n(z) \tag{2.3}$$

where the subscript n has been introduced to denote different eigensolutions. The spatially varying effective mass in the kinetic energy term couples the z and $x-y$ solutions. This is a severe complication, so some average effective mass m^* (typically the well effective mass) is used in that term instead. One may rewrite equation (2.3) as

$$\left[-\frac{\hbar^2}{2}\frac{d}{dz}\frac{1}{m^*(z)}\frac{d}{dz} + V(z) + \frac{\hbar^2 k^2}{2m^*}\left(\frac{m^*}{m^*(z)} - 1\right)\right] F_n(z) = \left(E_n\left(\overline{k}\right) - \frac{\hbar^2 k^2}{2m^*}\right) F_n(z) \tag{2.4}$$

If the well and barrier effective masses are similar, or if the kinetic energy is modest (typically true since QCLs tend to operate at low temperature), then the third term on the LHS is negligible. Under this simplification, equation (2.3) becomes

$$\left[-\frac{\hbar^2}{2}\frac{d}{dz}\frac{1}{m^*(z)}\frac{d}{dz} + V(z)\right] F_n(z) = E_n F_n(z) \tag{2.5}$$

where $E_n = E_n\left(\overline{k}\right) - \hbar^2 k^2/2m^*$ is the quantization energy associated with the $\hat{z}$ direction. Equation (2.5) reduces the QCL analysis to an essentially 1D problem, greatly easing analysis.

The final ingredient to completing the theoretical picture is the inclusion of boundary conditions. At the junction between two semiconductor layers 1 and 2, the envelope functions must satisfy

$$\begin{aligned} F_1 &= F_2 \\ \left(\frac{1}{m^*}\frac{dF}{dz}\right)_1 &= \left(\frac{1}{m^*}\frac{dF}{dz}\right)_2 \end{aligned} \tag{2.6}$$

As the effective masses in a superlattice are discontinuous, this manifest as a "kink" in the envelope function at layer boundaries.

2.2 Approaches to Transport: on the Choice of Basis

There are several formalisms for QCL transport using the envelope function theory of section 2.1. A detailed discussion of all of them is beyond the scope of this thesis, but the reader is the referred to the excellent review article by Wacker [28] for more information.

2.2.1 Miniband conduction

Except for the spatially varying effective mass, the form of superlattice envelope function equation (2.5) is exactly analogous to Schrödinger's equation for an electron in a periodic 1D potential. As such, one obvious approach to superlattice transport is to simply elevate all the standard techniques of solid-state physics to the level of the envelope functions. Although this picture is generally not used for QCL analysis, it is important from a conceptual viewpoint.

In this approach, one diagonalizes the Hamiltonian of the unbiased superlattice, assuming Bloch periodicity of the envelope function between superlattice periods. This results in a superlattice *miniband* dispersion, with the resulting *Wannier-Bloch* eigenstates analogous to the Bloch functions of a 1D atomic lattice (see figure 1-3). The applied electric field is treated as a perturbation inducing an electron in a given band to traverse the miniband Brillouin zone. Scattering processes induce changes between different k_z states and minibands.

Because the superlattice period is many hundreds of monolayers in length, the miniband Brillouin zone is much smaller than the microscopic Brillouin zone of bulk crystals. As such, it is much easier for an electron to fully traverse the miniband Brillouin zone. One proposal for original Esaki-Tsu superlattice was to exploit these *Bloch oscillations* for generating coherent THz radiation, and indeed, *Bloch gain* has recently been confirmed in a QCL [29].

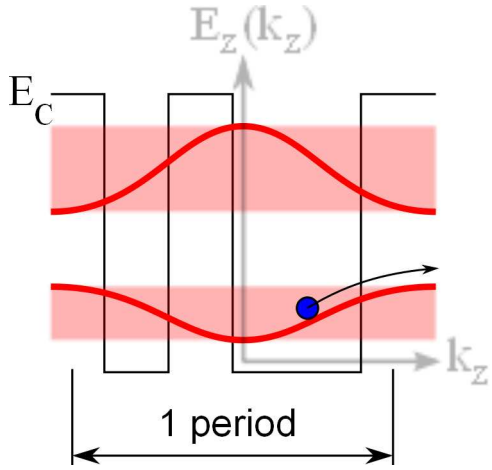


Figure 2-1: Schematic of miniband conduction. Only the superlattice potential is included in the Hamiltonian to be diagonalized, resulting in miniband continua of states indexed by a "supermomenta" k_z. Scattering happens between different minibands and k_z states, and an electric field induces evolution in k_z (motion in the super-Brillouin zone of envelope functions).

2.2.2 Wannier-Stark hopping: the semiclassical rate equations

This formalism chooses the eigenstates of the biased superlattice, the so-called *Wannier-Stark* states, as the basis. Assuming periodicity of the QCL potential, these states are periodic in energy and space: for any given eigenstate $\psi(z)$ with energy E, there will be a whole set of corresponding states with eigenfunctions $\psi(z-nd)$, and energies $E - e\mathcal{E}nd$, where $\mathcal{E}$ is is the applied electric field, d is the superlattice period, and n is an integer. This is known as the *Wannier-Stark ladder* (see figure 2-2), and is the usual basis for QCL calculations. Transport occurs through scattering between these eigenstates, in the manner of the classical Boltzmann transport equation. The scattering rates themselves, however, are calculated quantum mechanically using Fermi's golden rule (hence this approach is "semiclassical").

There are, however, some theoretical difficulties with the Wannier-Stark states; although not so relevant to practical design, they are worth mentioning. The application of an electric field yields (up to some additive constant) the potential $e\Phi(z) = -e\mathcal{E}z$. A QCL superlattice is typically hundreds of periods long, making it effectively infi-

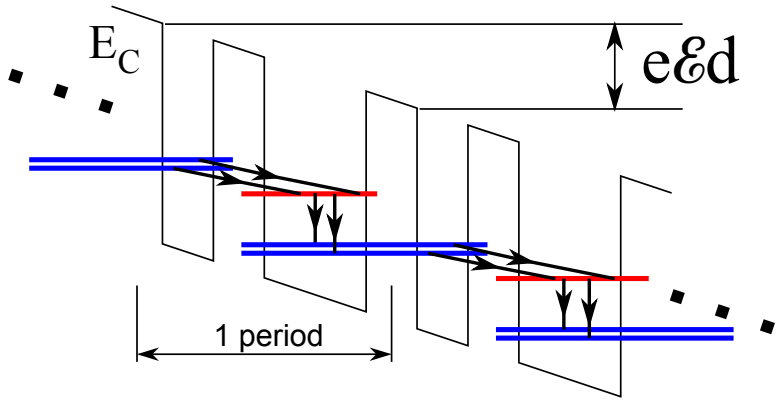


Figure 2-2: Schematic of Wannier-Stark Hopping conduction. Both the superlattice potential and the electric field are diagonalized, and transport occurs entirely through scattering between eigenstates. Equivalent WS states in adjacent periods are separated in energy by $e\mathcal{E}d$.

nite in length. This results in a potential that extends essentially from $-\infty$ to ∞ in energy, and consequently there are *no* bound states in the superlattice. In fact, the spectrum of the biased superlattice is mathematically proven to be continuous [30], even for arbitrarily small biases. On the other hand, the spectrum of an unbiased superlattice clearly has bound states, the experimental gain spectra of QCLs clearly shows discrete structure, and experimental absorption studies in other superlattice structures also support the existence of the Wannier-Stark ladder.

The resolution to this apparent paradox is that lasing in QCLs occurs between *resonant* eigenstates that possess a discrete spectrum embedded inside a continuous spectrum. One way to think about these resonant Wannier-Stark states is that they are the bound states of the unbiased superlattice, but the applied field induces them to possess a finite lifetime. This reflects that in the presence of a field, electrons possess a finite probability of leaking out of the quantum wells over time. As the field approaches zero, the escape rate approaches zero, or equivalently, the lifetime goes to infinity.

But the wavefunctions of the Wannier-Stark states are proven to have exponentially divergent tails at infinity, and hence are not unnormalizable in the usual sense ($\langle\psi|\psi\rangle = 1$). This reflects that in the absence of scattering, electrons in a resonant state will eventually leak out, and then accelerate indefinitely [31]. Fermi's golden

rule cannot be used with such pathological wavefunctions. Here, QCL design takes a refreshingly straightforward approach: for want of a better solution, the issue of unbounded wavefunctions is simply ignored. One or two modules of the superlattice are modeled as a closed system with Dirichlet boundary conditions ($F(z) = 0$ at simulation boundaries), and the resulting wavefunctions are normalized over the one or two module lengths. The resulting wavefunctions are well behaved, and Fermi's golden rule applies.

There is no mathematically rigorous justification for this artificial approach to boundary conditions. Physically, however, the heavy scattering in semiconductor systems would prevent the indefinite acceleration corresponding to the unnormalizable tail of the Wannier-Stark states. The states of interest are quite often located energetically far below adjacent quantum barriers, such that one would expect the resonant lifetime to be long in comparison to scattering induced lifetimes. In this situation, the expectation that the electrons should be localized to within one or two modules is reasonable. However, the resonant lifetimes grow shorter at high applied fields, and the lifetime associated with "over-barrier" escaping electrons may possibly become significant even in comparison to scattering mechanisms.

Perhaps the strongest endorsement of this method, however, is purely empirical: practically all QCLs are designed using this approach, and the success of all these devices attest to the validity of this method, even though its theoretical foundations may not be solid.

Analytical QCL rate equation model

As an example of a rate equation approach, QCLs are often described by a restricted rate equation model examining only the upper and lower laser subbands coupled to the optical field in the device. Although this model attains reasonable quantitative agreement with mid-IR devices, it is insufficient for THz devices. Nevertheless, it deserves a brief overview here on account of its ubiquity in the literature, and also because it has some utility in the qualitative description of THz QCLs.

Consider a QCL consisting of N_m modules and volume V enclosed in a waveguide

whose modes have uniform photon density (unachievable in practice of course, but consideration of the detailed modal information garners no additional insight). Let the upper and lower laser subband populations in the *total* device volume be denoted as n_3 and n_2 respectively (the designation "1" is normally reserved for the injector subband; see chapter 3). Let there be a single lasing mode of negligible spectral width, modal volume $V_{\rm ph}$ and total photon number $n_{\rm ph}$. Then the rate equations for n_3, n_2 and $n_{\rm ph}$ are

$$\begin{aligned}
\frac{dn_3}{dt} &= \frac{\eta I N_m}{e} - \frac{n_3}{\tau_{\rm sp}} - \frac{n_3 - n_2}{\tau'_{\rm sp}} \left(\frac{n_{\rm ph}}{V_{\rm ph}} \cdot \frac{V}{N_m} \right) - \frac{n_3}{\tau_3} \\
\frac{dn_2}{dt} &= \frac{(1-\eta) I N_m}{e} + \frac{n_3}{\tau_{\rm sp}} + \frac{n_3 - n_2}{\tau'_{\rm sp}} \left(\frac{n_{\rm ph}}{V_{\rm ph}} \cdot \frac{V}{N_m} \right) + \frac{n_3}{\tau_{32}} - \frac{n_2}{\tau_2} \\
\frac{dn_{\rm ph}}{dt} &= \frac{n_3}{\tau'_{\rm sp}} + \frac{n_3 - n_2}{\tau'_{\rm sp}} \left(\frac{n_{\rm ph}}{V_{\rm ph}} \cdot \frac{V}{N_m} \right) - \frac{n_{\rm ph}}{\tau_{\rm ph}}
\end{aligned} \tag{2.7}$$

The other symbols above are defined as follows:

- I is the device current. It is taken to be a model parameter.
- $0 < \eta < 1$ is the injection efficiency into the upper level. It is taken to be a model parameter.
- $\tau_{\rm sp}$ is the total spontaneous emission rate, and $\tau'_{\rm sp}$ is the spontaneous emission rate into the lasing mode.
- τ_{32} is the scattering lifetime of non-radiative processes from the upper to lower subband.
- τ_3 is the scattering lifetime of other upper level non-radiative processes.
- τ_2 is the scattering lifetime of lower level non-radiative processes.
- $\tau_{\rm ph}$ is the photon lifetime in the optical cavity.

The term $\left(\frac{n_{\rm ph}}{V_{\rm ph}} \cdot \frac{V}{N_m} \right)$ reflects that stimulated emission is proportional to the number of photons in each module, rather than the total number of photons. One may

abbreviate V/V_{ph} as Γ, the *confinement factor*, which measures the fraction of the optical mode overlapping with the active region.

In steady state, one sets all time derivatives in equations (2.7) to zero, and solves for the electron and photon populations. Above threshold, one must be careful to solve the equations with the amount of population inversion $n_3 - n_2$ held constant. Of particular interest is the degree of subthreshold population inversion, as this determines whether there will be positive gain for lasing. Subthreshold radiative lifetimes are typically orders of magnitude higher than non-radiative lifetimes, so one may neglect all radiative terms. The result is that

$$\Delta n = n_3 - n_2 = \frac{IN_m}{e}\left[\eta\tau_3\left(1 - \frac{\tau_2}{\tau_{32}}\right) - (1-\eta)\tau_2\right] \tag{2.8}$$

Let $\eta \sim 1$ (this is not necessarily—or even likely to be— true, but is done for simplicity due to the difficulty of determining η from experiment or calculation). One finds that

$$\Delta n \propto \tau_3\left(1 - \frac{\tau_2}{\tau_{32}}\right) \tag{2.9}$$

This simple rate equation model can be developed much further; the reader is referred to [32] for more elaborate treatments. However, due to the model's limited applicability to THz devices, this section stops here.

2.2.3 Density matrix transport: the Kazarinov-Suris model

As a Boltzmann transport equation-like model, the Wannier-Stark hopping model described in section 2.2.2 admits a simple physical description and appeals to intuition. Current conduction occurs between the Wannier-Stark eigenstates having different average positions, enabling an electron to hop down the superlattice. However, further examination reveals a number of contradictions.

From a conceptual viewpoint, whereas Wannier-Stark hopping describes current as being driven by scattering, the miniband conduction model describes current as caused by coherence. In the miniband conduction model of transport, scattering induces transitions to different k_z states, possibly in different minibands, but does not

change the average position of the electrons. Instead, the field-induced coherent evolution of the Wannier-Bloch states carries the current, and scattering is only relevant insofar as it interrupts the evolution.

From a empirical viewpoint, Callebaut has shown that the Wannier-Stark hopping model predicts unphysical behavior for current densities [33]. This is best explained by an example. Figure 2-3a depicts a two-well superlattice biased such that two single well states align in energy as indicated. When this *resonance*[1] or *anticrossing* occurs, diagonalization to calculate the Wannier-Stark states yields a delocalized bonding/anti-bonding pair separated by an *anticrossing gap*, Δ. In the Wannier-Stark hopping model, transport occurs through scattering from this pair into subsequent modules. Although increasing the barrier thickness decreases the size of the anticrossing gap, the wavefunctions are nearly unchanged, resulting in a current density essentially independent of barrier width. Experimentally, increasing the barrier thickness does decrease current density, a result consistent with physical intuition.

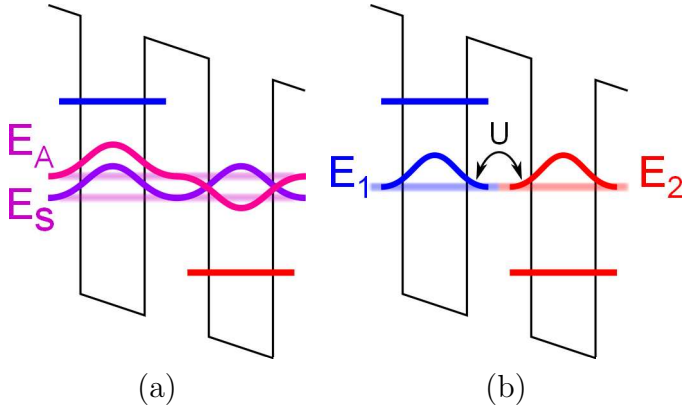


Figure 2-3: (a) Extended wavefunctions used in Wannier-Stark hopping. (b) Localized wavefunctions coupled through a barrier used in the Kazarinov-Suris model. The gap between the extended states is $\Delta = E_A - E_S \approx 2U$

The unphysical behavior of the Wannier-Stark hopping model is due to the neglect of coherent transport. To account for coherent effects, Kazarinov and Suris proposed a phenomenological density matrix model based on localized states (such as single well states[2]). Consider the same two-well superlattice in figure 2-3b. This time, the

[1]Not to be confused with the "resonant eigenstates" discussed in section 2.2.2; those are eigenstates with finite lifetime even in the absence of scattering.

[2]These have some complications in that they are not strictly orthogonal, which the present

two states in resonance are modeled as a dissipative two-level system coupled through an interaction potential U. In matrix form, the Hamiltonian is

$$H = \begin{bmatrix} E_1 & -U \\ -U & E_2 \end{bmatrix} \tag{2.10}$$

Diagonalizing H yields bonding and anti-bonding Wannier-Stark levels as before, with the coupling U responsible for the anticrossing gap at resonance ($E_1 = E_2$). The density matrix evolves according to

$$i\hbar \frac{d}{dt}\rho = [H, \rho] + i\hbar\Gamma(\rho) \tag{2.11}$$

where the commutator $[H, \rho]$ describes coherent evolution and Γ is the dissipation super-operator describing coupling to the environment (in this case, other superlattice states). Γ is approximated using a phenomenological relaxation time expression of the form

$$\Gamma(\rho) = \begin{bmatrix} \frac{\rho_{22}}{\tau} & -\frac{\rho_{12}}{\tau_{||}} \\ -\frac{\rho_{21}}{\tau_{||}} & -\frac{\rho_{22}}{\tau} \end{bmatrix} \tag{2.12}$$

In this picture, transport occurs via two mechanisms. A particle initially localized in state 1 in the left well will resonantly tunnel into state 2, and back again, much like the Rabi oscillations of an optical two-level system; these oscillations are interrupted (damped) by the scattering of electrons from state 2 into state 1 of a subsequent superlattice period, with some time constant τ. In addition, there may be processes that do not change populations (the diagonal elements ρ_{11} and ρ_{22}), but nonetheless scramble the coherence of the tunneling process; the overall phase-breaking time $\tau_{||}$ combines the effects of these "pure" dephasing processes with the phase-breaking due to population changing scattering. Let the total electron population be N, such that $\rho_{11} + \rho_{22} = N$. Solving equation (2.11) for steady state yields the Kazarinov-Suris

discussion ignores. A more appropriate choice might be, say, the Wannier function basis. The crucial point is that the basis is localized.

expression for the current through the barrier,

$$J = \frac{eN\rho_{22}}{\tau} = eN\frac{2\Omega^2\tau_{||}}{4\Omega^2\tau\tau_{||} + \omega_{21}^2\tau_{||}^2 + 1} \tag{2.13}$$

where $\Omega = U/\hbar$ and $\omega_{21} = (E_2 - E_1)/\hbar$.

Consider the case of resonance ($\omega_{21} = 0$). For strongly interacting states (thin barrier), $|\Omega| \gg 1/\sqrt{\tau\tau_{||}}$. Electrons in this regime oscillate freely between states 1 and 2 with only light damping due to scattering effects (oscillations are underdamped). This case yields

$$J \approx \frac{eN}{2\tau} \tag{2.14}$$

This expression is indeed independent of the interaction Ω between the two states and the phase-breaking time $\tau_{||}$. Transport is determined essentially by the scattering time in this regime, so one expects the Wannier-Stark hopping model to be valid. In contrast, for weakly interacting states (thick barrier), $|\Omega| \ll 1/\sqrt{\tau\tau_{||}}$. Electrons in this regime scatter rapidly but tunnel slowly (oscillations are overdamped). This case yields

$$J \approx 2eN\Omega^2\tau_{||} \tag{2.15}$$

Here one finds that current density is strongly dependent on the interaction between the two states and is sensitive to phase-breaking processes. This recovers the experimental result that current density does depend on the barrier width, as the barrier width factors into both Ω and $\tau_{||}$.

Note that the relaxation time approximation for Γ does *not* recover sensible results if the delocalized Wannier-Stark states are used as the basis. Although physical results should in principle be independent of the quantum basis, equation (2.12) for Γ is approximate, and our choice of basis affects the quality of the approximation. This can be illustrated by an example. At resonance ($E_1 = E_2$), diagonalizing equation (2.10) yields

$$H = SH_DS^\dagger \tag{2.16}$$

where

$$H_D = \begin{bmatrix} E-U & 0 \\ 0 & E+U \end{bmatrix} = \begin{bmatrix} E_S & 0 \\ 0 & E_A \end{bmatrix}, \; S = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{2.17}$$

The density matrix equation of motion (2.11) can be changed to the eigenbasis by multiplying both sides from the left by $S^\dagger$ and from the right by S (in this particular example $S = S^\dagger$). The result is

$$i\hbar \frac{d}{dt}\rho_D = [H_D, \rho_D] + i\hbar\Gamma_D(\rho_D) \tag{2.18}$$

where

$$\begin{aligned} \rho_D &= S^\dagger \rho S = \begin{bmatrix} \rho_{SS} & \rho_{SA} \\ \rho_{AS} & \rho_{AA} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2}(\rho_{11}+\rho_{12}+\rho_{21}+\rho_{22}) & \frac{1}{2}(\rho_{11}-\rho_{12}+\rho_{21}-\rho_{22}) \\ \frac{1}{2}(\rho_{11}+\rho_{12}-\rho_{21}-\rho_{22}) & \frac{1}{2}(\rho_{11}-\rho_{12}-\rho_{21}+\rho_{22}) \end{bmatrix} \end{aligned} \tag{2.19}$$

$$\begin{aligned} \rho &= S\rho_D S^\dagger = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2}(\rho_{SS}+\rho_{AS}+\rho_{SA}+\rho_{AA}) & \frac{1}{2}(\rho_{SS}-\rho_{AS}+\rho_{SA}-\rho_{AA}) \\ \frac{1}{2}(\rho_{SS}+\rho_{AS}-\rho_{SA}-\rho_{AA}) & \frac{1}{2}(\rho_{SS}-\rho_{AS}-\rho_{SA}+\rho_{AA}) \end{bmatrix} \end{aligned} \tag{2.20}$$

and

$$\begin{aligned} &\Gamma_D = S^\dagger \Gamma S \\ &= \begin{bmatrix} -\frac{1}{2\tau_{||}}(\rho_{21}+\rho_{12}) & \frac{\rho_{22}}{\tau} - \frac{1}{2\tau_{||}}(\rho_{21}-\rho_{12}) \\ \frac{\rho_{22}}{\tau} + \frac{1}{2\tau_{||}}(\rho_{21}-\rho_{12}) & \frac{1}{2\tau_{||}}(\rho_{21}+\rho_{12}) \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{2\tau_{||}}(\rho_{SS}-\rho_{AA}) & \frac{1}{2\tau}(\rho_{SS}-\rho_{AS}-\rho_{SA}+\rho_{AA}) \\ & -\frac{1}{2\tau_{||}}(\rho_{AS}-\rho_{SA}) \\ \frac{1}{2\tau}(\rho_{SS}-\rho_{AS}-\rho_{SA}+\rho_{AA}) & \\ +\frac{1}{2\tau_{||}}(\rho_{AS}-\rho_{SA}) & \frac{1}{2\tau_{||}}(\rho_{SS}-\rho_{AA}) \end{bmatrix} \end{aligned} \tag{2.21}$$

The on-diagonal scattering has the expected form of a rate equation involving the populations, but the off-diagonals do not evolve through a simple relaxation of the form $-\rho_{AS}/\tau_{D||}$.

Use of the Wannier-Stark basis leads to the erroneous conclusion that coherence plays no role in QCL transport. Unfortunately, Iotti and Rossi in their 2001 paper [34] introducing the use of ensemble Monte Carlo methods for QCL transport appears to have done exactly this: they employed a density matrix model in the Wannier-Stark basis with a scattering superoperator analogous in form to equation (2.12), with rates calculated using Fermi's golden rule. Perhaps this is the reason why they came to the conclusion that a QCL's transport is essentially incoherent. Callebaut modified the Monte Carlo density matrix technique to employ a tight-binding basis [35], and in doing so recovered the importance of coherence to transport. Weber et al. have also studied QCLs using a first principles density matrix formalism in the Wannier-Stark basis, concluding that coherence is crucial to transport [36]. The importance of coherence in QCL transport has also been affirmed in studies using the methods of NEGF [26, 37].

Quantitative application of the Kazarinov-Suris model is difficult because of large uncertainty in τ and $\tau_{||}$, and because often more than two states are of interest. However, it highlights the key role of the anticrossing gap Δ of delocalized Wannier-Stark states. This is because the interaction between the corresponding localized states can be estimated by $U \approx \Delta/2$, and hence Δ is a key metric of resonant tunneling through thick barriers, where Wannier-Stark hopping is invalid. Chapter 3 considers in detail the *injection anticrossing* and the *collection anticrossing*.

2.3 Numerical Determination of Eigenfunctions

Equation (2.5) admits tractable analytical solutions only for a few simple potentials, such as the prototypical unbiased, single quantum well. In general, a numerical solution is faster and more fruitful. This section covers two common methods, and a useful, but lesser known third.

2.3.1 Finite difference method

The finite difference method approximates the differential equation to be solved by its values on a mesh of uniformly spaced points. Let the computational domain be of length L, meshed with N points spaced by $\delta z = L/(N-1)$, and assume Dirichlet boundary conditions. A 2nd order finite difference approximation of the envelope function derivative is

$$\frac{d}{dz}F(z) \approx \frac{F(z+\delta z) - F(z-\delta z)}{2\delta z} \tag{2.22}$$

Interestingly, if $\underline{F}$ is a vector whose entries are the values of the envelope function at the grid points ($F[i] = F(i \times \delta z)$), then equation (2.22) defines a linear matrix algebra transformation for $\underline{F}$. That is, if $\underline{G}$ is the vector representing the derivative of $F(z)$, then $\underline{F}$ and $\underline{G}$ are related by

$$\underline{G} = \underline{\underline{D}}_1 \underline{F} \tag{2.23}$$

where $\underline{\underline{D}}_1$ is the infinite, tridiagonal matrix

$$\underline{\underline{D}}_1 = \frac{1}{2\delta z}\begin{bmatrix} \ddots & \ddots & & & \\ \ddots & 0 & 1 & & \\ & -1 & 0 & 1 & \\ & & -1 & 0 & \ddots \\ & & & \ddots & \ddots \end{bmatrix} \tag{2.24}$$

Similarly, multiplication by a scalar function, such as the potential $V(z)$, can be represented as multiplication of $\underline{F}$ by a diagonal matrix.

$$\underline{\underline{V}} = \begin{bmatrix} \ddots & & & & \\ & V[0] & & & \\ & & V[1] & & \\ & & & V[2] & \\ & & & & \ddots \end{bmatrix} \tag{2.25}$$

where, like the definition for $\underline{F}$, $V[i] = V(i \times \delta z)$. Therefore, one may write down by inspection the finite difference approximation of equation (2.5) as a matrix eigenvalue problem.

$$\left[-\frac{\hbar^2}{2} \underline{\underline{D}}_1 \underline{\underline{M}}^{-1} \underline{\underline{D}}_1 + \underline{\underline{V}} \right] \underline{F} = E \underline{F} \tag{2.26}$$

where $\underline{\underline{M}}^{-1}$ is the inverse effective mass matrix, which is diagonal like $\underline{\underline{V}}$. Finally, enforcing Dirichlet boundary conditions means $F[0] = 0$ and $F[N] = 0$; this enables truncation of the infinite dimensional matrix value problem of equation (2.26) into a $(N-2) \times (N-2)$ matrix value problem, which can be solved by any eigenvalue method of numerical linear algebra.

2.3.2 Shooting method

The finite difference methods above returns $N-2$ eigenstates, but one is normally only interested in a few of the lowest energy eigenstates. The numerical approximation worsens as the characteristic oscillations of the envelope function approaches the finite difference step size, so beyond some point in energy, the computed eigenstates degenerate into numerical garbage. This motivates the use of the shooting method, which directly computes the lowest energy eigenstates one at a time.

The shooting method for eigenfunction determination solves a 1D boundary value problem using the numerical methods of time-dependent ordinary differential equations. A typical time-dependent problem starts at an initial condition at some time t, and uses some approximation of the differential equation to determine the solution at a later time $t' = t + \delta t$. In this manner, the solution can be traced in time by steps of δt.[3] In the shooting method, the same procedure is applied to a spatial coordinate. The solution at some boundary z_1 is integrated forward by steps of δz until a second boundary z_2 is reached. However, in the shooting method, some of the boundary con-

[3]Although the section assumes constant δt, this is not necessary in general. Clever shooting method implementations may adjust δt in response to how quickly they perceive the solution to be changing at a given current point in time, optimizing the trade-off between resolution and solution speed.

ditions at z_1 and the eigenvalues are left as adjustable parameters to fit the boundary conditions at z_2. This motivates the name "shooting method": one targets ("shoots at") a boundary condition on the opposite side of the computational domain.

An example is helpful for clarification; the following shooting method is presented by Harrison for the solution of equation (2.5) [38]. One starts by deriving the following finite difference approximation of equation 2.5:

$$\begin{aligned}
&\left[-\frac{\hbar^2}{2}\frac{d}{dz}\frac{1}{m^*(z)}\frac{d}{dz}+V(z)\right]F(z)=EF(z)\\
\Rightarrow\ &\frac{-\hbar^2}{2\delta z}\left[\frac{1}{m^*(z+\delta z/2)}\left.\frac{dF}{dz}\right|_{z+\delta z/2}-\frac{1}{m^*(z-\delta z/2)}\left.\frac{dF}{dz}\right|_{z-\delta z/2}\right]+V(z)\\
&=E_nF_n(z)\\
\Rightarrow\ &\frac{-\hbar^2}{2\delta z}\left[\frac{1}{m^*(z+\delta z/2)}\frac{F(z+\delta z)-F(z)}{\delta z}-\frac{1}{m^*(z-\delta z/2)}\frac{F(z)-F(z-\delta z)}{\delta z}\right]\\
&+V(z)=E_nF_n(z)\\
\Rightarrow\ &\frac{-\hbar^2}{2(\delta z)^2}\left[\frac{F(z+\delta z)-F(z)}{m^*(z+\delta z/2)}-\frac{F(z)-F(z-\delta z)}{m^*(z-\delta z/2)}\right]+V(z)=EF(z) \qquad (2.27)
\end{aligned}$$

In the above, Harrison takes $m^*(z\pm\delta z/2)$ to be the average effective mass between two points, but another possible choice is to instead calculate the average inverse effective mass between two points; the difference is small unless the heterostructure has extremely mismatched effective masses. One may rewrite equation (2.27) to yield the solution for a subsequent point as function of the previous two points. This reads

$$\begin{aligned}
\frac{F(z+\delta z)}{m^*(z+\delta z/2)} &= \left[\frac{2(\delta z)^2}{\hbar^2}(V(z)-E)+\frac{1}{m^*(z+\delta z/2)}+\frac{1}{m^*(z-\delta z/2)}\right]F(z)\\
&\quad -\frac{F(z-\delta z)}{m^*(z-\delta z/2)} \qquad (2.28)
\end{aligned}$$

Using Dirichlet boundary conditions, the initial conditions at the left-hand side of the simulation boundary are that $F(z=0)=0$ and $F(z=\delta z)$ is any non-zero value. The arbitrariness of $F(z=\delta z)$ is because eigenfunctions are unique only up to a multiplicative constant, said constant being determined by wavefunction normalization requirements and itself determined only up to a complex phase factor.

Integrating over to $z = L$, one treats $F(z = L)$ as a function of E, and scans E through some range of energies to find solutions satisfying $F(z = L) = 0$. A root-finding algorithm (for example, the secant method or Newton-Ralphson method) is used to refine the value of E when close to an eigenvalue.

Paul and Fouckhardt introduced a superior shooting method in [39]. In this alternate method, one defines

$$G(z) = \frac{1}{m^*(z)} \frac{d}{dz} F(z) \tag{2.29}$$

and specifies different, but coupled, shooting equations for F and G. This yields

$$\begin{aligned} G(z + \delta z) &= G(z) + \frac{2\delta z}{\hbar^2}(V(z) - E)F(z) \\ F(z + \delta z) &= F(z) + \delta z \; m^*(z)G(z) \end{aligned} \tag{2.30}$$

These equations are simpler to compute than equation (2.28), and reference [39] also shows them to be more numerically stable.

2.3.3 Spectral element method

Before discussing the spectral element method, one must first understand the more specific technique of *spectral methods*. This thesis could not possibly do justice to this rich subject area, so the introduction here is necessarily qualitative. The interested reader is referred to Boyd's excellent textbook [40].

Spectral methods take a philosophically different approach to numerical solution from the finite difference approximation on which both the finite difference method and the shooting method is based. Spectral methods pick some basis of orthogonal functions, and assume that the solution of a problem can be expanded in this basis. One then attempts to find the expansion coefficients of the series. Whereas accuracy of a solution is determined by the step size δz in the finite difference approximation, it is determined by the number of terms retained in the expansion in spectral methods. Qualitatively, one could say that finite difference method finds exact solutions to

an approximation of the *problem*, whereas spectral methods try to find approximate *solutions* to the exact problem. A well-chosen basis results in *exponentially* fast convergence. Since finite difference methods converge only algebraically, spectral element methods practically always achieve better accuracy for a given number of points—sometimes dramatically so.

Like in quantum mechanics, the choice of basis depends somewhat on the problem at hand, but is also largely a matter of taste. Surprisingly, there exist a class of orthogonal functions whose expansion coefficients are *equal* to the value of the solution at certain points (nodes), the *Lagrange interpolation polynomials*. It turns out that if these points are chosen to be zeros or extrema of another basis set (usually the Chebyshev polynomials of the first kind), the Lagrange interpolation polynomials will inherit the same convergence properties. Ironically, the practical implication of using the Lagrange interpolation polynomials as a spectral basis is that the spectral method is merely solving for the values of the solution on a non-uniform grid. Phrased as such, this *pseudospectral* method does not sound so different from the finite difference method, despite the philosophical differences explained previously.

A major difficulty with spectral methods, however, is that they lose their excellent convergence properties in the vicinity of discontinuities. A well-known example of this is Gibb's phenomena seen in the Fourier expansion of piecewise continuous functions. Unfortunately, the envelope function problem for a heterostructure is riddled with discontinuities, both in the effective mass and in the potential. The solution, as first advocated by Patera [41], is to divide the computational domain into smooth sub-domains, or *elements*, each possessing its own series expansion, and then enforce boundary conditions between elements. A superlattice lends itself well to this approach, as one can treat each layer as a separate element. This essentially combines spectral methods with the finite element method, resulting in the *spectral element method*.

Unfortunately, the mathematics of the spectral element method are significantly more complicated to derive. The pseudospectral variant used in this thesis is based on the method reported in [42]. It uses the *Gauss-Legendre-Lobatto* (GLL) nodes, and

possesses the convergence properties of a direct expansion in Legendre polynomials. On the normalized interval [-1,1], the N GLL nodes are the extrema of the $(N-2)$th Legendre polynomial and the two end-points ± 1. Denoting these as ζ_n, they can be remapped as necessary onto any other linear interval $[a, b]$ through the transformation.

$$z_n = \frac{b+a}{2} + \zeta_n \frac{b-a}{2} \tag{2.31}$$

Let the Lagrange interpolation polynomials be denoted $b_n(z)$. The derivation starts by considering just a single element. First, the inner produce (integral) of both sides of equation (2.5) is taken against an arbitrary $b_m(z)$ (readers familiar with finite element methods will recognize this as Galerkin's method). This yields

$$-\frac{\hbar^2}{2} \int_C dz \; b_m \frac{d}{dz} \frac{1}{m^*} \frac{d}{dz} F + \int_C dz b_m V F = E \int_C dz b_m F \tag{2.32}$$

where C is the length of the element under consideration, and the argument dependences of the functions have been suppressed for clarity (eg. F really should be $F(z)$). The first term is expanded using integration by parts

$$\begin{aligned} & \int_C dz \; b_m \frac{d}{dz} \frac{1}{m^*} \frac{d}{dz} F \\ = & \left[\frac{b_m}{m^*} \frac{dF}{dz} \right]_{\partial C} - \int_C dz \; \frac{db_m}{dz} \frac{1}{m^*} \frac{dF}{dz} \\ = & \; \delta_{m0} \left[\frac{1}{m^*} \frac{dF}{dz} \right]_{z=z_0} - \delta_{mN} \left[\frac{1}{m^*} \frac{dF}{dz} \right]_{z=z_N} - \int_C dz \; \frac{db_m}{dz} \frac{1}{m^*} \frac{dF}{dz} \end{aligned} \tag{2.33}$$

The last step above uses a property of the Lagrange interpolation polynomials that $b_m(z_n) = \delta_{mn}$, where z_n is the nth GLL node. Next, the series expansion

$$F(z) = \sum_{n=1}^{N} F_n b_n(z), \tag{2.34}$$

where $F_n = F(z_n)$, is inserted into equations (2.32) and (2.33), yielding

$$-\delta_{m0}\left[\frac{\hbar^2}{2m^*}\frac{dF}{dz}\right]_{z=z_0} + \delta_{mN}\left[\frac{\hbar^2}{2m^*}\frac{dF}{dz}\right]_{z=z_N} + \frac{\hbar^2}{2}\sum_{n=1}^{N}\left[\int_C dz\ \frac{db_m}{dz}\frac{1}{m^*}\frac{db_n}{dz}\right]F_n$$
$$+\sum_{n=1}^{N}\left[\int_C dz\ V b_m b_n\right]F_n = E\sum_{n=1}^{N}\left[\int_C dz b_m b_n\right]F_n \tag{2.35}$$

In principle, the integrals above can be integrated analytically, but numerical integration in this case is much simpler. Conveniently, the GLL nodes are the same nodes at which a function needs to be evaluated in order to perform *Gaussian quadrature*. Gaussian quadrature of order N approximates a normalized integral in [-1,1] as

$$\int_{-1}^{1} f(\zeta)d\zeta \approx \sum_{p=1}^{N} w_p f(\zeta_p) \tag{2.36}$$

and for an arbitrary interval $[a, b]$,

$$\int_a^b f(z)dz \approx \sum_{n=1}^{N}\frac{(b-a)w_p}{2}f(z_p) = \sum_{p=1}^{N} W_p f_p. \tag{2.37}$$

(The integration weights w_p are given in [40]). Applying Gaussian quadrature and after doing some algebra, equation (2.35) becomes

$$-\delta_{m0}\left[\frac{\hbar^2}{2m^*}\frac{dF}{dz}\right]_{z=z_0} + \delta_{mN}\left[\frac{\hbar^2}{2m^*}\frac{dF}{dz}\right]_{z=z_N} + \frac{\hbar^2}{2}\sum_{n=1}^{N}\left[\sum_{p=1}^{N}\frac{db_m(z_p)}{dz}\frac{W_p}{m_p^*}\frac{db_n(z_p)}{dz}\right]F_n$$
$$+V_m W_m F_m = E W_m F_m \tag{2.38}$$

Equation (2.38) defines a set of N equations in the element of consideration; essentially, it leaves one with a mini-matrix problem for each element, and it remains to combine them. Consider now two adjacent regions A and B, with N_A and N_B points respectively. At the boundary between A and B, boundary conditions (2.6) imply that $F_A(z_{N,A}) = F_B(z_{0,B})$, motivating one to add together the last equation of A and the first equation of B. But the boundary terms $\left[\frac{\hbar^2}{2m^*}\frac{dF_A}{dz}\right]_{z=z_{N,A}=z_{0,B}}$ and

$\left[\frac{\hbar^2}{2m^*}\frac{dF_B}{dz}\right]_{z=z_{N,A}=z_{0,B}}$ are equal, and so cancel. The essential point is that all the internal boundary terms between elements cancel out when the sets of equations for different elements are combined. Furthermore, Dirichlet boundary conditions enables one to discard the external boundary terms at the edges of the computational domain. Ultimately, one finds that equation (2.5) can be approximated by the generalized, linear, matrix eigenvalue problem

$$\left[\frac{\hbar^2}{2}\underline{\underline{D}}_1^T\underline{\underline{M}}^{-1}\underline{\underline{W}}\;\underline{\underline{D}}_1+\underline{\underline{V}}\;\underline{\underline{W}}\right]\underline{F}=\underline{\underline{A}}\;\underline{F}=E\underline{\underline{W}}\;\underline{F} \tag{2.39}$$

It is difficult to discuss the structure of the matrices without a specific example, so consider a computational domain partitioned into 3 elements with 4, 3 and 5 nodes, respectively. Consider $\underline{\underline{D}}_1$: each element will contribute an elemental sub-matrix and these will be assembled into $\underline{\underline{D}}_1$ by "overlapping and adding together at the corners." For example, the elemental sub-matrix of the first element has elements defined by

$$\left[\underline{\underline{D}}_1^{(1)}\right]_{mn}=\frac{db_n^{(1)}(z_m)}{dz} \tag{2.40}$$

where $b_n^{(1)}$ are the Legendre interpolation polynomials for region 1 (formulae for the pseudospectral differentiation matrices are given in [40, 43]). $\underline{\underline{D}}_1$ is schematically shown in 2-4. The inverse effective mass matrix $\underline{\underline{M}}^{-1}$ and the matrix of Gaussian quadrature weights $\underline{\underline{W}}$ are similarly assembled from the inverse masses and quadrature weights evaluated at the nodes in each element. (Conveniently, these last two matrices are diagonal.)

The problem specification is complete at this point, but Cheng et al. [42] introduce a final numerical method to turn generalized eigenvalue problem (2.39) into a regular eigenvalue problem by manipulating equation (2.39) into

$$\begin{aligned}
&\underline{\underline{A}}\;\underline{F}=E\underline{\underline{W}}\;\underline{F}\\
\Rightarrow\;&\underline{\underline{W}}^{-1/2}\underline{\underline{A}}\left(\underline{\underline{W}}^{-1/2}\underline{\underline{W}}^{1/2}\right)\underline{F}=E\underline{\underline{W}}^{1/2}\;\underline{F}\\
\Rightarrow\;&\underline{\underline{A}}'\underline{F}'=E\underline{F}'
\end{aligned} \tag{2.41}$$

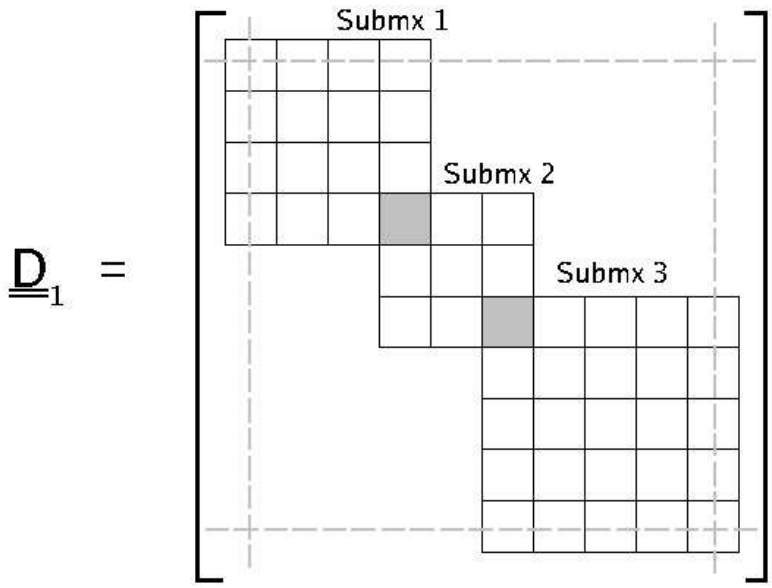


Figure 2-4: Schematic of pseudospectral differentiation matrix for the spectral element method, for a hypothetical computational domain partitioned into three elements having 4, 3, and 5 nodes, respectively. The differentiation matrices for each region are assembled into $\underline{\underline{D}}_1$ by overlapping and adding together boundary entries (indicated in grey). Dirichlet boundary conditions enable one to strike out the first and last rows and columns (dotted lines).

Where $\underline{\underline{A}}' = \underline{\underline{W}}^{-1/2} \underline{\underline{A}}\, \underline{\underline{W}}^{-1/2}$ and $\underline{F}' = \underline{\underline{W}}^{1/2} \underline{F}$. As with the finite difference method, this eigenvalue problem can now be solved using any eigensolver.

2.3.4 Comparison of methods

Both the finite difference method and spectral element method require linear algebra eigensolvers; the most basic and reliable of these is the QR method, and its cousin for generalized eigenvalue problems, the QZ method [40]. Unfortunately, the QR/QZ methods can be numerically expensive, with costs scaling as $O(N^3)$ where N is the matrix size or number of nodes (these being approximately equal). In contrast, assuming some average number of root-finding iterations for each eigenfunction, the shooting method scales as $O(NM)$, where M is the number of points to scan in energy (assuming linear scan). M will depend on the number of eigenfunctions requested and the nature of the particular potential. For example, one might expect that M would scale as the square of the number of requested eigenfunctions in an

infinite quantum well. But in general, M will be modest compared to N^2. This favorable scaling makes the shooting method potentially the fastest of the three methods, achieving finite difference accuracy at a fraction of the cost of eigensolver routines when N becomes large.

Unfortunately, the shooting method's speed comes at the cost of stability. From quantum mechanics, Schrödinger's equation in regions where the potential energy surpasses the particle energy leads to exponentially growing or decaying waves with imaginary wavevector. The exponentially divergent solutions frequently lead to numerical overflow when one integrates from a well into a thick barrier, causing the shooting method to crash (for this reason it is also necessary to shoot in the direction of decreasing electric potential energy). Although the shooting method usually does converges, it is the author's experience that there is no easy way of predicting when it does not, nor of salvaging broken problems. In contrast, both the finite difference method and spectral element method are very robust.

The advantage of spectral elements over finite differences is that the exponential accuracy of the former enables one to use far fewer solution nodes. This greatly speeds up the QR algorithm. As numerical methods are not the focus of this thesis, this section makes no attempt at a controlled study comparing finite differences and spectral elements (see [42] for such a study), but nevertheless, a semi-quantitative example is illuminating.

This thesis uses a custom spectral element code developed in MATLAB 7.0. The tool of choice for the last three generations of graduate students working on the THz QCL project at MIT is the SEQUAL 2.1 finite difference code developed in the 1980s at Purdue University [44] (the order of the finite difference approximation is unknown). The spectral element code of this thesis normally uses 1 node every 2 monolayers, whereas the SEQUAL 2.1 convention is to use 2 nodes per monolayer. Ignoring prefactors to the computational costs, the QR algorithm scaling suggests a 64-fold reduction in computation time. To test this, table 2.1 compares the performance of this thesis's code and SEQUAL 2.1 in computing the two-module eigenstates of QCL design OWI185E-M1 (see section 4.2.3). It is difficult to get both programs to

	SEM			SEQUAL 2.1		
E_{ij}	0.5pts /ML (meV)	1.0pts /ML (meV)	Rel. err (%)	2.0pts /ML (meV)	4.0pts /ML (meV)	Rel.err (%)
E_{32}	37.80	37.80	0.466e-4	37.90	37.94	-0.0933
E_{43}	5.138	5.138	5.765e-4	5.127	5.129	-0.0544
E_{54}	5.137	5.137	-2.628e-4	5.162	5.164	-0.0341
E_{65}	13.75	13.75	-0.327e-4	13.95	13.98	-0.2419
E_{76}	3.212	3.212	11.65e-4	3.182	3.176	0.2009
E_{87}	35.57	35.57	-1.435e-4	35.72	35.76	-0.1157
E_{98}	4.809	4.809	5.868e-4	4.802	4.8	0.03521
$E_{10,9}$	2.737	2.737	-15.10e-4	2.257	2.238	0.8526
$E_{11,10}$	2.621	2.621	13.48e-4	3.133	3.16	-0.8472
$E_{12,11}$	9.425	9.425	-3.812e-4	9.138	9.095	0.4791
$E_{13,12}$	7.192	7.192	13.23e-4	7.627	7.685	-0.7600
$E_{14,13}$	23.78	23.78	-4.075e-4	23.33	23.28	0.1980
$E_{15,14}$	13.72	13.72	0.010e-04	13.81	13.83	-0.1357
Time(s)	0.083	0.600		2.75	23.0	

Table 2.1: Performance comparison of custom spectral element method (SEM) implemented in MATLAB 7.0 versus SEQUAL 2.1. Reported are the energy differences between adjacent eigenstates at different resolutions. The relative error is specified relative to the higher resolution calculation for both methods.

agree on exactly the same eigenvalues (due to slight differences in parameters, rounding, etc.), so table 2.1 compares relative accuracy for each program: for each method, the eigenvalue problem is solved, and then the calculation is repeated with twice the resolution. The results from the differing resolutions are compared to see how well converged the solution is. Although the results of table 2.1 are not well controlled (for example, SEQUAL must spend spend additional time writing results to hard drive, making the contest stacked against it), the spectral element method clearly outperforms SEQUAL 2.1 in both speed and relative accuracy. The actual speed increases are $\sim$33 between the two low resolution computations, and $\sim$38 between the two high resolution calculations. The reasons for the factor of $\sim$2 discrepancy from 64 are not known.

A few minor points in closing: there is also a risk of the shooting method accidentally skipping over closely spaced eigenvalues, but it is the author's experience that the aforementioned stability problems are a more severe issue. On the other hand,

the shooting method can account for non-parabolicity using a energy-dependent effective mass at no additional cost. In contrast, equation 2.5 with an energy dependent effective mass cannot be cast into a linear eigenvalue problem, so the finite difference method and spectral element method both require iterative solutions where none were necessary before. However, non-parabolicity is generally not considered in this thesis. Finally, the spectral element method's non-uniform grid accommodates layers of arbitrary thickness, not just multiples of the step-size δz as in the finite difference and shooting methods.

2.4 Non-radiative Scattering Mechanisms

Having established the eigenstates of the biased superlattice, this section considers the scattering mechanisms which drive transport. All scattering rates here are calculated using Fermi's golden rule.[4] The treatment here is standard. The derivations below are also covered in Smet [45], Callebaut [33], Harrison [38], and other excellent sources.

The eigenstates of the superlattice are uniquely identified by a subband number and an in-plane momentum.[5] In this context, Fermi's golden rule for scattering from a state $i, \overline{k}_i$ (subband i with momentum $\overline{k}_i$, energy $E_i(\overline{k}_i) = E_i + \frac{\hbar^2 k_i^2}{2m^*}$, and envelope function $F_i(z)\frac{e^{i\overline{k}_i \cdot \overline{r}}}{\sqrt{A}}$) to a state $f, \overline{k}_f$ (subband f with momentum $\overline{k}_f$, energy $E_f(\overline{k}_f) = E_f + \frac{\hbar^2 k_f^2}{2m^*}$, envelope function $F_f(z)\frac{e^{i\overline{k}_f \cdot \overline{r}}}{\sqrt{A}}$) due to a time-harmonic perturbing potential $H'(t) = H'e^{\pm i\Delta E/\hbar t}$ reads

$$W\left(i, \overline{k}_i \rightarrow f, \overline{k}_f\right) = \frac{2\pi}{\hbar} \left|\left\langle f, \overline{k}_f \mid H' \mid i, \overline{k}_i \right\rangle\right|^2 \delta\left(E_f(\overline{k}_f) - E_i(\overline{k}_i) \pm \Delta E\right) \quad (2.42)$$

where ΔE is the energy absorbed from or released to an intermediate entity such as a phonon or photon, and δ represents the *Dirac delta function*; its presence reflects that energy must be conserved over long time scales.

A note on the calculation of matrix elements in Fermi's Golden Rule: given two

[4] First formulated by Dirac, Fermi being its titular spokesman notwithstanding.

[5] Spin is ignored here because spin-up and spin-down are degenerate in the absence of an applied magnetic field

perturbations H_1' and H_2', one normally proceeds to calculate $\left|\left\langle f, \overline{k}_f \mid H_1' \mid i, \overline{k}_i \right\rangle\right|^2$ and $\left|\left\langle f, \overline{k}_f \mid H_2' \mid i, \overline{k}_i \right\rangle\right|^2$ instead of $\left|\left\langle f, \overline{k}_f \mid H_1' + H_2' \mid i, \overline{k}_i \right\rangle\right|^2$. While the latter matrix element seems more mathematically well motivated, the cross-terms between H_1' and H_2' complicates analysis. The standard approach is much more convenient because, by neglecting these cross-terms (correlations), scattering rates for each mechanism can be calculated independently and directly added together. The neglect of cross-terms is physically equivalent to assuming uncorrelated scattering mechanisms, which is an assumption somewhat justifiable given that Fermi's golden rule is accurate only in reflecting mean behavior of weak scattering over long time scales. But ultimately, the real motivation is probably mathematical convenience: given that any golden rule analysis is only approximate anyway, the extra rigor likely nets little advantage over the simpler approach.

2.4.1 Scattering by longitudinal optical phonons

Scattering by longitudinal optical (LO) phonons is the most important scattering mechanism in QCL transport, since it yields sub-picosecond scattering times in polar semiconductors such as GaAs/AlGaAs. This is a two-edged sword. The phonon depopulation QCL designs studied in this thesis (see chapter 3) use this ultrafast LO phonon scattering for depopulating the lower laser subband to create population inversion, but it is also hypothesized to be a major source of non-radiative scattering of electrons from the upper laser subband (recall section 1.1).

The section treats scattering of subband electrons from a thermalized gas of dispersionless, bulk GaAs LO phonon. This treatment is approximate, since the differing mechanical properties of quantum wells and quantum barriers will induce quantization in the phonon spectrum as well. However, studies by Williams [46] and Gao [47] suggest that calculations based on detailed phonon spectra make little difference to overall scattering rates. Because QCLs are mostly GaAs (the barriers are typically only 15%-30% aluminum), even in more detailed calculations scattering tends to be dominated by GaAs-like phonon modes [46]. Moreover, screening and renormalization

effects due to charges from other electrons are ignored.

Let the phonon momentum be denoted $\overline{Q}$ ($\overline{q}$ will be reserved for the in-plane momentum). The interaction Hamiltonian is

$$H'_{LO}(t) = \sum_{\overline{Q}} \alpha(\overline{Q}) \left[b(\overline{Q},t) e^{i\overline{Q}\cdot\overline{r}} + b^{\dagger}(\overline{Q},t) e^{-i\overline{Q}\cdot\overline{r}} \right] \tag{2.43}$$

where $b^{\dagger}(\overline{Q},t) = b^{\dagger}_{\overline{Q}} e^{-i\omega_{LO}t}$ and $b(\overline{Q},t) = b_{\overline{Q}} e^{i\omega_{LO}t}$ are phonon creation and annihilation operators. The term $\alpha\left(\overline{Q}\right)$ is the *Frölich interaction strength* given by

$$\left|\alpha(\overline{Q})\right|^2 = \frac{1}{V} \frac{e^2 \hbar \omega_{LO}}{2Q^2} \left(\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \tag{2.44}$$

where V is the system volume, $\hbar\omega_{LO} \approx 36$meV is the LO phonon energy of GaAs, and $\epsilon_{\infty} = 10.89\epsilon_0$ and $\epsilon_s = 12.19\epsilon_0$ are the optical and static permittivities of bulk GaAs. The matrix element is therefore

$$\begin{aligned} & \left|\left\langle f, \overline{k}_f \mid H'_{LO} \mid i, \overline{k}_i \right\rangle\right|^2 \\ = \; & \frac{1}{V} \sum_{\overline{Q}} \frac{e^2 \hbar \omega_{LO}}{2} \left(\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \frac{\left|A_{if}(q_z)\right|^2}{q_z^2 + q^2} \delta_{\overline{k}_f - \overline{k}_i = \mp\overline{q}} \left(N_{LO} + \frac{1}{2} \pm \frac{1}{2} \right) \\ = \; & \frac{e^2 \hbar \omega_{LO}}{2A} \left(\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \left(N_{LO} + \frac{1}{2} \pm \frac{1}{2} \right) \int \frac{dq_z}{2\pi} \frac{\left|A_{if}(q_z)\right|^2}{q_z^2 + q^2} \end{aligned} \tag{2.45}$$

where $q^2 = \left|\overline{k}_f - \overline{k}_i\right|^2$, N_{LO} is the Bose-Einstein occupation factor for phonons at a lattice temperature T_L

$$N_{LO} = \frac{1}{\exp\left(\frac{\hbar\omega_{LO}}{k_B T_L}\right) - 1} \tag{2.46}$$

and A_{if} is dubbed the LO phonon scattering form factor

$$A_{if}(q_z) = \int_{-\infty}^{\infty} dz \psi_f^*(z) \psi_i(z) e^{\pm i q_z z} \tag{2.47}$$

In equations (2.45) and (2.47), the upper-sign corresponds to phonon emission, and the lower sign corresponds to phonon absorption. The derivation of equation (2.45) uses standard sum-to-integral substitutions for $\lim V \to \infty$: $\frac{1}{V}\sum_{\overline{Q}} \to \int \frac{d^3\overline{Q}}{2\pi}$, $\delta_{\overline{k}_f - \overline{k}_i = \mp\overline{q}} \to$

$\frac{(2\pi)^2}{A}\delta\left(\overline{k}_f - \overline{k}_i \pm \overline{q}\right)$. Fermi's golden rule reads

$$\begin{aligned} & W\left(i, \overline{k}_i \rightarrow f, \overline{k}_f\right) \\ = & \frac{2\pi}{\hbar}\frac{e^2\hbar\omega_{LO}}{2A}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_s}\right)\left(N_{LO} + \frac{1}{2} \pm \frac{1}{2}\right) \\ & \times\left(\int \frac{dq_z}{2\pi}\frac{|A_{if}(q_z)|^2}{q_z^2 + q^2}\right)\delta\left(E_f(\overline{k}_f) - E_i(\overline{k}_i) \pm \hbar\omega_{LO}\right) \end{aligned} \tag{2.48}$$

In practice, one is generally interested in an average scattering rate between two subbands, rather than between single particle states. This requires some integrals over the initial and final momenta. The first step is to sum over all possible final momenta (state blocking is neglected here). The required integral is tedious to compute, but final result is

$$\begin{aligned} & W(i, \overline{k}_i \rightarrow f) \\ = & \Theta\left[\pm\left(E_i + \frac{\hbar^2 k_i^2}{2m^*} - E_f\right) > \hbar\omega_{LO}\right] \\ & \times C_{LO}\left(N_{LO} + \frac{1}{2} \pm \frac{1}{2}\right)\int_{-\infty}^{\infty} dq_z \frac{|A_{if}(q_z)|^2}{\sqrt{q_z^4 + 2q_z^2(2k_i^2 - \kappa^2) + \kappa^4}} \end{aligned} \tag{2.49}$$

where

$$\kappa^2 = \frac{2m^*}{\hbar^2}(E_f - E_i \pm \hbar\omega_{LO}) \tag{2.50}$$

and the scattering prefactor C_{LO} is given by

$$\begin{aligned} C_{LO} &= \frac{1}{4}\left(\frac{m^*}{\pi\hbar^2}\right)\hbar\omega_{LO}\left(\frac{e^2}{\hbar}\right)\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_s}\right) \\ &= \alpha_{fs}\left(\frac{m^* c}{\hbar^2}\right)\hbar\omega_{LO}\left(\frac{\epsilon_0}{\epsilon_\infty} - \frac{\epsilon_0}{\epsilon_s}\right) \end{aligned} \tag{2.51}$$

where α_{fs} is the fine structure constant. The leading step function (Θ) in equation (2.49) expresses the fact that some transitions are energetically forbidden. For example, no LO emission is possible if the initial state is below the minimum energy of the final subband.

Finally, the scattering rate is *averaged* (not summed!) over the initial momentum

distribution, to find the mean carrier lifetime. If the subband i has momentum distribution $f_i(\overline{k})$, then the average scattering rate is

$$\left\langle W(i,\overline{k}_i \to f)\right\rangle_{\overline{k}_i} = \frac{\sum_{\overline{k}} f_i(\overline{k}) W(i,\overline{k} \to f)}{\sum_{\overline{k}} f_i(\overline{k})} \tag{2.52}$$

It is usual to assume thermalized subbands, such that the subband distribution function is a Fermi-Dirac distribution dependent only on kinetic energy; that is, $f_i(\overline{k}) = f_i(E_k)$. Also, equation (2.49) shows that the scattering rate is also a function only of kinetic energy, ie. $W(i,\overline{k} \to f) = W_{if}(E_k)$. Therefore, one may write

$$\left\langle W(i,\overline{k}_i \to f)\right\rangle_{\overline{k}_i} = \frac{\int_0^\infty dE_k\ g(E_k) f_i(E_k) W_{if}(E_k)}{\int_0^\infty dE_k\ g(E_k) f_i(E_k)} = \frac{\int_0^\infty dE_k f_i(E_k) W_{if}(E_k)}{\int_0^\infty dE_k\ f_i(E_k)} \tag{2.53}$$

where $g(E_k)$ is the subband density of states; within the approximation of constant effective mass, $g(E_k)$ is conveniently constant for 2D subbands, and therefore cancels out between the numerator and denominator. But instead of performing the full thermal average, for $E_f - E_i < \hbar\omega_{LO}$, a useful approximation is that

$$\left\langle W(i\overline{k}_i \to f)\right\rangle_{\overline{k}_i} = \frac{1}{\tau_{if}} \approx W_{if}^{\text{hot}} \exp\left(\frac{E_f - E_i - \hbar\omega_{LO}}{k_B T_L}\right) \tag{2.54}$$

where W_{if}^{hot} is the scattering rate at the lowest E_k in the subband where LO-phonon scattering is energetically allowed [46]. QCL design and analysis frequently employs equation (2.54).

A final note: equation (2.49) is not in the usual form encountered in the literature (Harrison is the exception here [38]). Most literature sources replace the integral in equation (2.49) with the equivalent expression

$$\frac{1}{2}\int_0^{2\pi} d\theta B_{if}(q) \tag{2.55}$$

where

$$B_{if}(q) = \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dz' F_f^*(z) F_i(z) F_i^*(z') F_f(z') \frac{1}{q} e^{-q|z-z'|} \tag{2.56}$$

$$q^2 = \left|\overline{k}_f - \overline{k}_i\right|^2 = k_f^2 + k_i^2 - 2k_i k_f \cos\theta \tag{2.57}$$

The popularity of this more common expression is somewhat puzzling. Equation (2.55) requires evaluation of a triple integral, whereas equation (2.49) requires only a double integral. Equation (2.49) seems to have the clear advantage for computation.

2.5 Radiative Scattering

2.5.1 Fermi's golden rule for the light-matter interaction

This section considers radiative processes based on Fermi's golden rule. Although sophisticated theories of optical gain employing Green's functions exist [25], this simpler theory will suffice for the present.

Let the optical field be determined by a scalar potential $\phi(\overline{r}, t)$ and a vector potential $\overline{A}(\overline{r}, t)$. The usual *Coulomb gauge* is adopted here, where $\nabla \cdot \overline{A} = 0$ and $\phi = 0$. The *minimal-coupling* light-matter Hamiltonian is then obtained by replacing the canonical momentum $\overline{p}$ of the field-free Hamiltonian with the kinetic momentum $\overline{p} - e\overline{A}$, where e is the magnitude of the electron charge (the Hamiltonian of the field itself is ignored, as the quantum mechanics of the light field are not of interest). However, envelope function equation (2.1) is *not* Schrödinger's equation, and the effective mass Hamiltonian is not amenable to this treatment. Assuming a homogeneous semiconductor for the moment, Schrödinger's equation in the presence of an optical field is

$$H\Psi = \left[\frac{(\overline{p} - e\overline{A})^2}{2m_0} + V_L\right]\Psi = i\hbar\frac{\partial}{\partial t}\Psi \tag{2.58}$$

where V_L is the microscopic lattice potential and m_0 is the bare electron mass. The

Hamiltonian expands as

$$\begin{aligned} & \frac{(\overline{p}-e\overline{A})^2}{2m_0}+V_L \\ = & \frac{p^2}{2m_0}+V_L-\frac{e}{2m_0}\left(\overline{p}\cdot\overline{A}+\overline{A}\cdot\overline{p}\right)+\frac{e^2A^2}{2m_0} \\ \approx & \left[\frac{p^2}{2m_0}+V_L\right]-\frac{e}{m_0}\overline{A}\cdot\overline{p} \end{aligned} \tag{2.59}$$

The above derivation employs the fact that $\overline{p}\cdot(\overline{A}f)=(\overline{p}\cdot\overline{A})f+\overline{A}\cdot(\overline{p}f)=\overline{A}\cdot(\overline{p}f)$ since $\overline{p}\cdot\overline{A}=-i\hbar\nabla\cdot\overline{A}=0$ for our choice of gauge. It also neglects the term quadratic in $\overline{A}$, as this is unimportant to linear optics. In equation (2.59), the term in square brackets is the Hamiltonian for a single electron in a periodic lattice. Approximating it with the effective mass Hamiltonian yields

$$\begin{aligned} & \left[\frac{p^2}{2m_0}+V_L\right]-\frac{e}{m_0}\overline{A}\cdot\overline{p} \\ = & \frac{p^2}{2m^*}-\frac{e}{m_0}\overline{A}\cdot\overline{p} \end{aligned} \tag{2.60}$$

Allowing for the effects of a heterostructure and an external potential, the Hamiltonian further changes to

$$\overline{p}\frac{1}{2m^*(\overline{r})}\overline{p}+V(\overline{r})-\frac{e}{m_0}\overline{A}\cdot\overline{p}=H_0+H'_{e-l} \tag{2.61}$$

The light-matter coupling Hamiltonian is therefore

$$H'_{e-l}=-\frac{e}{m_0}\overline{A}\cdot\overline{p} \tag{2.62}$$

To calculate matrix elements in Fermi's Golden rule, at this point we make the envelope function approximation, and assume a set of eigenstates of H_0 given approximately by $\Psi_n(\overline{r})=F_n(\overline{r})u(\overline{r})$, where $u(\overline{r})$ is the microscopically varying Γ-point Bloch function (assumed to be approximately the same for all constituent materials of the heterostructure), and the F_n are the mesoscopically varying, orthonormal envelope functions. Within a one-band approximation as done here, $u(\overline{r})$ is the

same for all wavefunctions of interest. Let the envelope functions be normalized as $\int d^3\overline{r}\ F_m^*(\overline{r})F_n(\overline{r}) = \delta_{mn}$, and the Bloch function be normalized as $\int d^3\overline{r}\ u^*(\overline{r})u(\overline{r}) = V$, where both integrals are over the entire system (volume V). $\delta_{m,n}$ is the Kronecker delta function. Therefore,

$$\begin{aligned} \left\langle F_f u \mid H'_{e-l} \mid F_i u \right\rangle &= -\int d^3\overline{r}\ F_f^* u^* \frac{e}{m_0}\overline{A}\cdot\overline{p}\,(F_i u) \\ &= -\int d^3\overline{r}\ F_f^* \frac{e}{m_0}\overline{A}\cdot(\overline{p}F_i)\,u^* u - \int d^3\overline{r}\ F_f^* F_i u^* \frac{e}{m_0}\overline{A}\cdot(\overline{p}u) \end{aligned} \tag{2.63}$$

Next, the envelope function approximation is used to separate the integration of the rapidly varying u from the slowly varying F_n. At THz frequencies, the optical wavelengths ($\sim$0.1mm) are much longer than the periodicity of u ($\sim$3Å), and therefore $\overline{A}$ may also be separated from the integrations of u. Equation 2.63 is then approximately given by

$$\begin{aligned} &\approx -\sum_n F_f^*(\overline{R}_n)\frac{e}{m_0}\overline{A}(\overline{R}_n)\cdot(\overline{p}F_i)|_{\overline{R}_n}\int_{\Delta_n} d^3\overline{r}\ u^*(\overline{r})u(\overline{r}) \\ &\quad -\sum_n F_f^*(\overline{R}_n)F_i(\overline{R}_n)\frac{e}{m_0}\overline{A}(\overline{R}_n)\cdot\int_{\Delta_n} d^3\overline{r}\ u^*(\overline{r})\,(\overline{p}u(\overline{r})) \\ &\approx -\int d^3\overline{r}\ F_f^*\frac{e}{m_0}\overline{A}\cdot(\overline{p}F_i) \end{aligned} \tag{2.64}$$

where the summations are over all unit cells of the semiconductor crystal (the n-th cell having position $\overline{R}_n$ and volume Δ_n). The second summation is dropped because the integral $\int d^3\overline{r}\ u^*\,(\overline{p}u) = -i\hbar\int d^3\overline{r}\ u^*\,(\nabla u)$ is zero due to the symmetry properties of u and ∇u. Therefore, the calculation of matrix elements for Fermi's golden rule in equation (2.63) ignores contributions from the Bloch function matrix elements and uses just the envelope functions. That is,

$$\left\langle F_f u \mid H'_{e-l} \mid F_i u \right\rangle \approx -\int d^3\overline{r}\ F_f^*\frac{e}{m_0}\overline{A}\cdot\overline{p}F_i \tag{2.65}$$

One is reminded in the above that the envelopes are uniquely indexed by a subband and in-plane momentum ($|F_n\rangle = \left|n\overline{k}_n\right\rangle$).

In the quantum theory of fields, the vector potential $\overline{A}$ can be written as

$$\overline{A}(\overline{r},t) = \sum_{\lambda} \sqrt{\frac{\hbar}{2\epsilon\omega_\lambda}} \left(b_\lambda(t)\overline{u}_\lambda(\overline{r}) + b_\lambda^\dagger(t)\overline{u}_\lambda^*(\overline{r}) \right) \tag{2.66}$$

where $b_\lambda(t) = b_\lambda e^{-i\omega_\lambda t}$ and $b_\lambda^\dagger(t) = b_\lambda^\dagger e^{i\omega_\lambda t}$ are the annihilation and creation operators for the mode indexed by λ. Evaluating equation (2.65) yields that

$$\begin{aligned} & \left| \left\langle f\overline{k}_f \mid H'_{e-l} \mid i\overline{k}_i \right\rangle \right|^2 \\ = & \sum_{\lambda} \frac{e^2\hbar}{2m_0^2\epsilon\omega_\lambda} \left[N_\lambda \delta_{N_{\lambda f}, N_{\lambda i}-1} \left| \left\langle f\overline{k}_f \mid \overline{u}_\lambda \cdot \overline{p} \mid i\overline{k}_i \right\rangle \right|^2 \right. \\ & \left. + (N_\lambda + 1)\delta_{N_{\lambda f}, N_{\lambda i}+1} \left| \left\langle f\overline{k}_f \mid \overline{u}_\lambda^* \cdot \overline{p} \mid i\overline{k}_i \right\rangle \right|^2 \right] \end{aligned} \tag{2.67}$$

where $N_{\lambda f}$ and $N_{\lambda i}$ are the number of photons in the final and initial states respectively. The Kronecker deltas preserving photon number are hereafter ignored, as tracking the quantum statistics of the photons in the cavity is of no interest.

To proceed further, it is convenient to specialize to the case of free space, in which the modes are normalized plane waves indexed by a polarization σ and a wavevector $\overline{\beta}$, or

$$\overline{u}_\lambda = \overline{u}_{\sigma\overline{\beta}} = \frac{e^{i\overline{\beta}\cdot\overline{r}}}{\sqrt{V}} \hat{e}_{\sigma\overline{\beta}}. \tag{2.68}$$

where $\hat{e}_{\sigma\overline{\beta}}$ is a polarization vector. Then

$$\begin{aligned} & \left\langle f\overline{k}_f \mid \overline{u}_\lambda \cdot \overline{p} \mid i\overline{k}_i \right\rangle \\ = & \frac{1}{\sqrt{V}} \left\langle f\overline{k}_f \mid e^{i\overline{\beta}\cdot\overline{r}} \hat{e}_{\sigma\overline{\beta}} \cdot \overline{p} \mid i\overline{k}_i \right\rangle \approx \frac{1}{\sqrt{V}} \left\langle f\overline{k}_f \mid \hat{e}_{\sigma\overline{\beta}} \cdot \overline{p} \mid i\overline{k}_i \right\rangle \end{aligned} \tag{2.69}$$

The last step above employs the *electric dipole approximation*, taking $e^{i\overline{\beta}\cdot\overline{r}} \approx 1$. This is justified because the wavefunctions have an extent on the order of one module length at most ($\sim$500Å), whereas a THz wavelength is $\sim$0.1mm. An equivalent interpretation is that the photon momentum is negligible. Next, one inserts the relation $\overline{p} = \frac{im_0}{\hbar}[H_0, \overline{r}]$ (derivable from the canonical commutation relation $[\overline{p}, \overline{r}] = -i\hbar$) to

yield

$$
\begin{aligned}
& \frac{1}{\sqrt{V}} \left\langle f, \overline{k}_f \mid \hat{e}_{\sigma\overline{\beta}} \cdot \overline{p} \mid i, \overline{k}_i \right\rangle \\
= \; & \frac{im_0}{\hbar\sqrt{V}} \left\langle f, \overline{k}_f \mid \hat{e}_{\sigma\overline{\beta}} \cdot [H_0, \overline{r}] \mid i, \overline{k}_i \right\rangle \\
= \; & \frac{im_0}{\hbar\sqrt{V}} \left[E_f(\overline{k}_f) - E_i(\overline{k}_i) \right] \left\langle f, \overline{k}_f \mid \hat{e}_{\sigma\overline{\beta}} \cdot \overline{r} \mid i, \overline{k}_i \right\rangle \\
= \; & \frac{im_0}{\hbar\sqrt{V}} \left[E_f(\overline{k}_f) - E_i(\overline{k}_i) \right] \hat{e}_{\sigma\overline{\beta}} \cdot \left[\hat{z} \left\langle f \mid z \mid i \right\rangle \left\langle \overline{k}_f | \overline{k}_i \right\rangle + \langle f | i \rangle \left\langle \overline{k}_f \mid \overline{\rho} \mid \overline{k}_i \right\rangle \right] \\
= \; & \frac{im_0}{\hbar\sqrt{V}} \left[E_f - E_i \right] (\hat{e}_{\sigma\overline{\beta}} \cdot \hat{z}) z_{fi} \delta_{\overline{k}_f, \overline{k}_i}
\end{aligned} \tag{2.70}
$$

where z_{fi} is the z-projected *dipole matrix element*. Note that the term $(\hat{e}_{\sigma\overline{\beta}} \cdot \hat{z})$ forbids interactions with optical fields polarized in-plane.

Combining equations (2.70) and (2.67) and inserting into Fermi's golden rule (2.42) yields for the photon emission and absorption rates

$$
\begin{aligned}
& W_{\mathrm{em}}(i, \overline{k}_i \to f, \overline{k}_f) \\
= \; & \delta_{\overline{k}_f, \overline{k}_i} \sum_{\sigma\overline{\beta}} \left| \hat{e}_{\sigma\overline{\beta}} \cdot \hat{z} \right|^2 (n_{\sigma\overline{\beta}} + 1) \frac{\pi e^2 \omega_{\sigma\overline{\beta}}}{2\epsilon V} \left| z_{fi} \right|^2 \delta(E_f - E_i + \hbar\omega_{\sigma\overline{\beta}})
\end{aligned} \tag{2.71}
$$

$$
\begin{aligned}
& W_{\mathrm{abs}}(i, \overline{k}_i \to f, \overline{k}_f) \\
= \; & \delta_{\overline{k}_f, \overline{k}_i} \sum_{\sigma\overline{\beta}} \left| \hat{e}_{\sigma\overline{\beta}} \cdot \hat{z} \right|^2 n_{\sigma\overline{\beta}} \frac{\pi e^2 \omega_{\sigma\overline{\beta}}}{2\epsilon V} \left| z_{fi} \right|^2 \delta(E_f - E_i - \hbar\omega_{\sigma\overline{\beta}})
\end{aligned} \tag{2.72}
$$

In closing, note that the minimal-coupling momentum substitution applied directly to the effective mass Hamiltonian in equation (2.1) yields

$$
H'_{e-l} = -\frac{e}{m^*} \overline{A} \cdot \overline{p} \tag{2.73}
$$

for the light-matter interaction Hamiltonian, with only inner products with the envelope functions being relevant. The difference between equation (2.73) and the previously derived equation (2.62) is the use of the effective mass in place of the bare electron mass. Equation (2.73) appears frequently in the literature [46, 48], but appears to be incorrect, strictly speaking. But surprisingly, due to the cancellation

of the masses in substituting equation (2.70) into equation (2.67), both interaction Hamiltonians will recover the correct scattering rates.

2.5.2 Intersubband optical gain

The difference between the photon emission and absorption rates gives the rate at which photons are produced or absorbed—in other words, the optical gain or loss. Let there be just one free space optical mode possessing frequency ω and polarization parallel to the the growth axis, $\hat{z}$. The net radiative transitions between initial and final subbands i and f is given by

$$\begin{aligned} & W(i \to f) \\ = \; & 2 \sum_{\overline{k}_i} \sum_{\overline{k}_f} \left\{ W_{\mathrm{em}}(i, \overline{k}_i \to f, \overline{k}_f) f_i(\overline{k}_i) \left[1 - f_f(\overline{k}_f) \right] \right. \\ & \left. - W_{\mathrm{abs}}(f, \overline{k}_f \to i, \overline{k}_i) f_f(\overline{k}_f) \left[1 - f_i(\overline{k}_i) \right] \right\} \end{aligned} \tag{2.74}$$

where $f_n(\overline{k}_n)$ is the state occupation, and the factor of two accounts for spin-degeneracy. Inserting equations (2.71) and simplifying yields

$$\begin{aligned} & W(i \to f) \\ = \; & \frac{\pi e^2 \omega}{\epsilon V} \left| z_{fi} \right|^2 \delta(E_f - E_i + \hbar \omega_{\sigma \overline{\beta}}) \cdot \\ & \sum_{\overline{k}} N \left[f_i(\overline{k}) - f_f(\overline{k}) \right] + f_i(\overline{k}) \left[1 - f_f(\overline{k}) \right] \\ = \; & \frac{\pi e^2}{\epsilon V} \delta(E_f - E_i + \hbar \omega) \omega \left| z_{fi} \right|^2 \\ & \cdot \left[N(n_i - n_f) + \sum_{\overline{k}} f_i(\overline{k}) \left[1 - f_f(\overline{k}) \right] \right] \\ \approx \; & \frac{\pi e^2}{\epsilon V} \delta(E_f - E_i + \hbar \omega) \omega \left| z_{fi} \right|^2 N(n_i - n_f) \end{aligned} \tag{2.75}$$

where n_i and n_f are the 2-D subband populations (unitless), and N is the photon number. The last term in the above corresponds to spontaneous emission, which is presumed to be small in comparison to stimulated emission. In the above, the

δ-function can be taken out of the summation because the subband dispersions are parallel in the absence of nonparabolicity (joint density of states is a δ-function; all transitions are at the same frequency regardless of in-plane momentum).

Considering an incident photon flux of $(N/V)v = (N/V)c/n_r$, where n_r is the refractive index of the semiconductor, the gain is given by

$$\begin{aligned} g(\omega) &= \frac{W(i \to f)/V}{(N/V)c/n_r} \\ &= \frac{\pi e^2 n_r}{\epsilon c} \delta(E_f - E_i + \hbar\omega) \omega \left|z_{fi}\right|^2 \left(\frac{n_i - n_f}{V}\right) \end{aligned} \tag{2.76}$$

Energy level broadening (due to scattering induced level lifetimes) can be phenomenologically included by replacing the Dirac delta function in equation (2.76) with a normalized Lorentzian with a full-width at half-maximum (FWHM) of ΔE

$$\delta(E_f - E_i + \hbar\omega) \to \gamma(\hbar\omega) = \frac{\Delta E/2\pi}{(E_f - E_i + \hbar\omega)^2 + (\Delta E/2)^2} \tag{2.77}$$

to yield

$$g(\omega) = \frac{\pi e^2 n_r}{\epsilon c} \omega \left|z_{fi}\right|^2 \Delta N \gamma(\hbar\omega) \tag{2.78}$$

where $\Delta N = (n_i - n_f)/V$ is the population inversion per unit volume. Peak gain occurs when $\hbar\omega = h\nu_{if} = E_i - E_f$; the expression is

$$g_{\text{peak}} = \frac{2e^2 n_r}{\hbar \epsilon c} \left|z_{fi}\right|^2 \Delta N \frac{\nu_{if}}{\Delta\nu} \tag{2.79}$$

where $\Delta\nu$ is the linewidth in units of frequency.

While equation (2.79) is complete, it is useful to rewrite it in terms of another parameter, the *oscillator strength*, f_{fi}.

$$\begin{aligned} g_{\text{peak}} &= \frac{n_r e^2}{2\pi m^* \epsilon c} \frac{\Delta N}{\Delta\nu} f_{fi} \\ &\approx (70\text{cm}^{-1}) \frac{\Delta N/10^{15}\text{cm}^{-3}}{\Delta\nu/\text{THz}} f_{fi} \quad \text{(for GaAs)}, \end{aligned} \tag{2.80}$$

where

$$f_{fi} = \frac{2m^*(E_i - E_f)}{\hbar^2} |z_{fi}|^2 \tag{2.81}$$

Note from equation (2.79) that the peak gain does not have any explicit dependence on electron mass, effective or bare. Therefore, the choice of mass used in the light-matter Hamiltonian does not affect the final result.

2.5.3 The oscillator strength

The oscillator strength indicates the "strength" of an optical transition. It is analogous to the oscillator strength of classical electromagnetics, where the ideal Hertzian dipole is defined to have an oscillator strength of 1. As such, the gain of the transition between any two subbands levels is proportional to the oscillator strength. Due to the presence of the dipole matrix element, the oscillator strength is, in general, larger for extended states (such as at anticrossing) than for localized states. An optical transition with a lower oscillator strength is said to be more *diagonal*.

Any quantum system, *regardless* of the shape of the potential, has the property that the oscillator strengths between all possible transitions sum to unity.[6] Because of the formal equivalence between the envelope function equation (2.1) and Schrödinger's equation, this property carries over. For a system of constant effective mass, it can be shown that the oscillator strength obeys the sum rule defined by

$$\sum_{j \neq i} f_{ij} = 1 \tag{2.82}$$

This has interesting consequences in quantum systems. Because of the energy difference factor in the oscillator strength definition (equation 2.81), quantum oscillator strengths can be *negative*. This means that in the sum rule, oscillator strengths for other transitions can be *greater* than unity; in general, transitions between higher lying states will have larger oscillator strength.

When the effect of valence bands upon the conduction band is considered, one

[6] However, with a general quantum system, the oscillator strength is defined with the bare electron mass.

must consider non-parabolicity. The effective mass becomes energy dependent, and the envelope function equation (2.1) loses its formal equivalence to Schrödinger's equation. The sum rule equation (2.82) does not hold in this situation. Sirtori et al. discuss modifications to the sum rule due to non-parabolicity, and also spatially varying effective mass [49]. Moreover, one often encounters in the literature the oscillator strength defined with the bare electron mass rather than the effective mass in equation (2.81), in which case the sum rule is modified to

$$\sum_{j \neq i} f_{ij} = \frac{m^*}{m_0} \tag{2.83}$$

Oscillator strength is a crucial parameter in QCL design, and will be discussed at length in the discussion on temperature performance in the next chapter.

Combining equation (2.9) with the oscillator strength yields a useful figure-of-merit to which the gain is proportional.

$$g_{\text{peak}} \propto \text{FOM} = f_{fi}\tau_3 \left(1 - \frac{\tau_2}{\tau_{32}}\right) \tag{2.84}$$

Chapter 3

THz Quantum Cascade Laser Design

Having established the theoretical foundations of QCLs in Chapter 2, this chapter discusses design. Because of QCL periodicity, one need only specify the design of a single period (module).[1] A single period of a QCL superlattice can partitioned into an *active region* responsible for optical gain and an *injector region* responsible for funneling electrons into a subsequent module. This conceptual division exists even though these regions are not always physically distinct. The injector is also key to the electrical characteristics of a QCL, and a well designed injector is necessary to avoid operation in negative differential resistance (see section 3.2.3).

Different designs use different active and injection regions. At present, there are three major THz QCL designs

1. The *chirped superlattice* (CSL) design

2. The *bound-to-continuum design* (BTC) design

3. The *phonon depopulation design* (PD) design

The CSL design was introduced by Köhler et al. in the first THz QCL [16] (see

[1] This is a major simplifying assumption. Recent computational results call into question whether QCL operation is truly periodic (see [26], for example), but to assume otherwise dramatically complicates the problem.

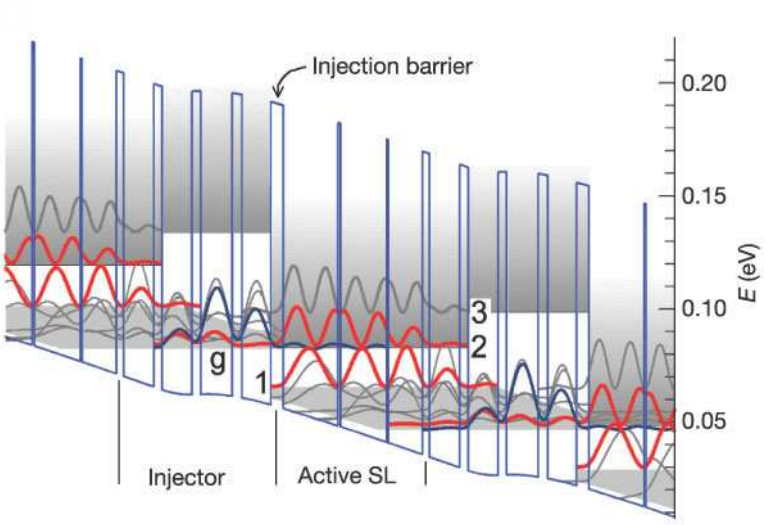
figure 3-1a). In this approach, the basic superlattice period consist of several wells of gradually narrowing width. At some bias the states of these wells align in energy to form highly anticrossed "minibands" of states. Electrons tend to relax to the bottom of the minibands through fast intersubband electron-electron scattering, thus population inversion can be achieved between an upper miniband and lower miniband of one module. The lower miniband is also designed to align with the upper miniband of a subsequent module, hence facilitating the cascading process through resonant tunneling.

The BTC design is a variant of the CSL design. One in the progression of narrowing wells is made slightly larger. When the biased to miniband formation, the aberrated well forms a "defect" state that lies in the gap between the upper and lower minibands (see 3-1b). Population inversion is formed between this defect state and the top of the lower miniband. The defect state is well localized ("bound") but emits to the highly delocalized ("continuum") lower miniband; this is the origin of the name "bound-to-continuum."[2] BTC designs are notable for their extremely low threshold current densities.

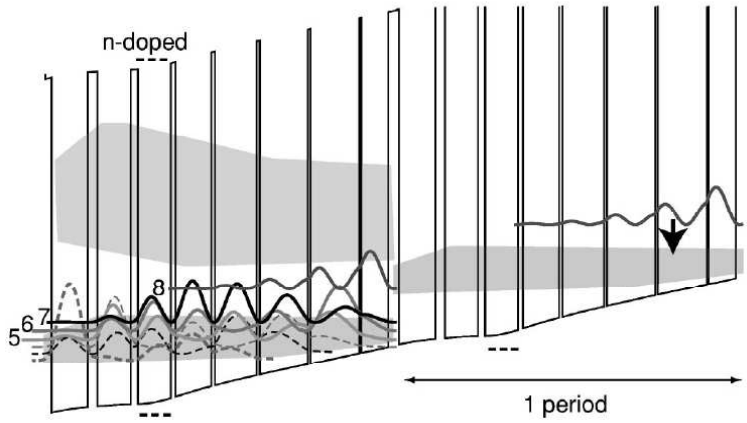
The PD design works on an entirely different principle from the CSL and BTC designs. There are two variants: *resonant-phonon* (RP) and *intrawell-phonon* (IP). Injection into the upper laser level is still accomplished using resonant tunneling, but using far fewer states. In the RP variant, the lower level is depopulated by resonant tunneling to an adjacent well followed by LO phonon emission. This form of phonon depopulation was used in the original mid-IR QCL [13], and adapted for use in THz by Williams et al. [51]. To date, RP based QCL designs possess the highest temperature performance. In the IP variant, there is no phonon depopulation step; the lower level is directly depopulated by LO phonon emission.

Although many of the concepts discussed in this chapter apply to QCLs in general, the remainder of this chapter focuses on PD designs, which are MIT's specialty. Note that the division into CSL, BTC, and PD is not strict. For example, there exist hybrid designs which combine BTC and RP features.

[2]This contrasts the CSL design, which could rightly be called a "continuum-to-continuum" design.



(a)



(b)

Figure 3-1: (a) Band diagram of CSL design by Köhler et al., reproduced from [16]. Population inversion occurs between level 2 at the bottom of the upper miniband and level 1 at top of the lower miniband. (b) Band diagram of BTC design by Scalari et al., reproduced from [50]. Injection occurs into the localized "defect" level 8. Population inversion occurs between level 8 and the top of the lower miniband.

3.1 Phonon Depopulation Device Families

This section discusses the major PD designs. A note on naming conventions: although there are no strict rules in this regard, MIT designs are generally labeled by a short acronym indicating particular features of the design, then the number of modules specified for growth, and finally by additional markings indicating different variations on the same design.

3.1.1 FL devices

The FL series contains the first of the operational resonant phonon designs. The acronym "FL" originally stood for "Four-well, LO-phonon depopulated," but the acronym has since lost its meaning since many members of this family have more than four quantum wells. This family of devices features active regions consisting of 2-3 wells and a injector region typically consisting of two wells. The band diagram for a sample FL design is shown and explained in figure 3-2.

FL designs traditionally have large oscillator strengths (very vertical transitions). This is largely for historical reasons: the FL series devices were based on the non-lasing L-series and M-series (see Xu [52] and Williams [46] for discussion of these devices). Prior to the achievement of lasing, research in RP THz QCLs was largely based on spontaneous emission measurements of the superlattices. Large oscillator strengths are required to avoid spontaneous emission measurements from being swamped by thermal background.

3.1.2 OWI devices

The acronym "OWI" stands for "One Well Injector." The OWI family are RP devices but the injector region consists of a single quantum well, and so there is only one subband injecting electrons into the subsequent module. The band diagram for a sample OWI design is shown and explained in figure 3-3.

The OWI devices were originally introduced for producing low frequency lasers [53]. Because the majority of electrons in a QCL reside in the injector region, a

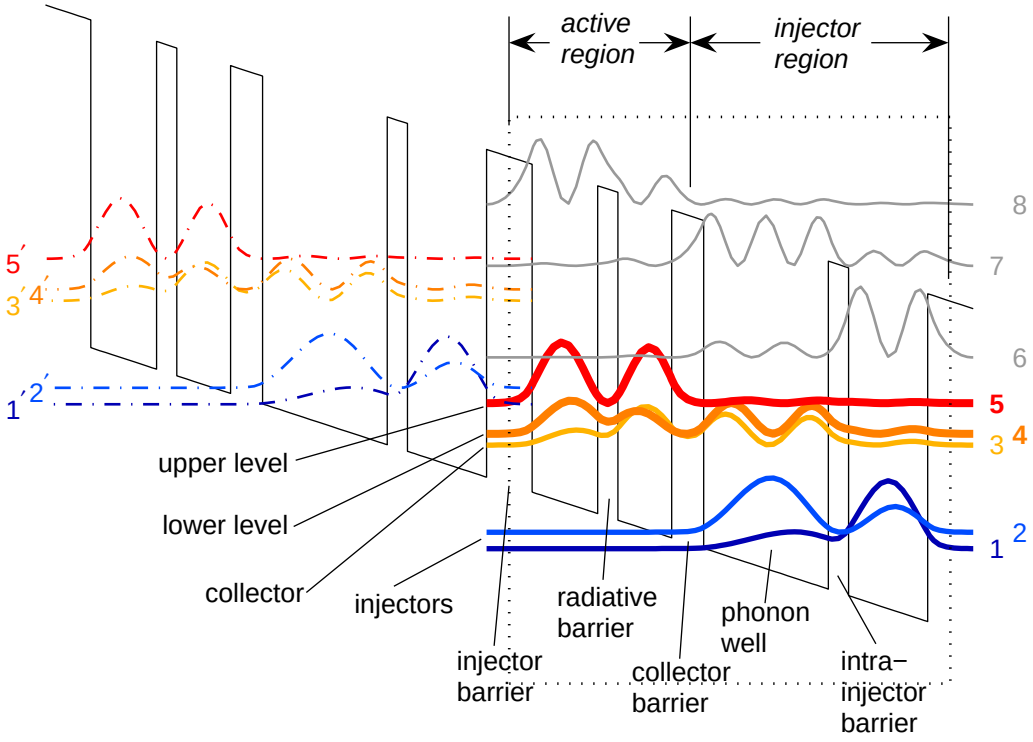


Figure 3-2: Sample band diagram of an FL family device (design FL175C from [46, 51]); one QCL period is boxed. At the design bias, the injector $1'$ of the previous module aligns in energy with the upper laser level ($1' - 5$ anticrossing), populating the upper level through resonant tunneling. After photon emission ($E_{54} = 13.9\text{meV} \equiv 3.4\text{THz}$), the lower level is depopulated by resonant tunneling extraction by the collector, which is the excited state of an adjacent well. E_{32} is 39.3meV, approximately resonant with the GaAs LO phonon energy of $E_{LO} \approx 36\text{meV}$. This in turn enables depopulation of the collector via fast LO phonon scattering into the injector levels 1 and 2.

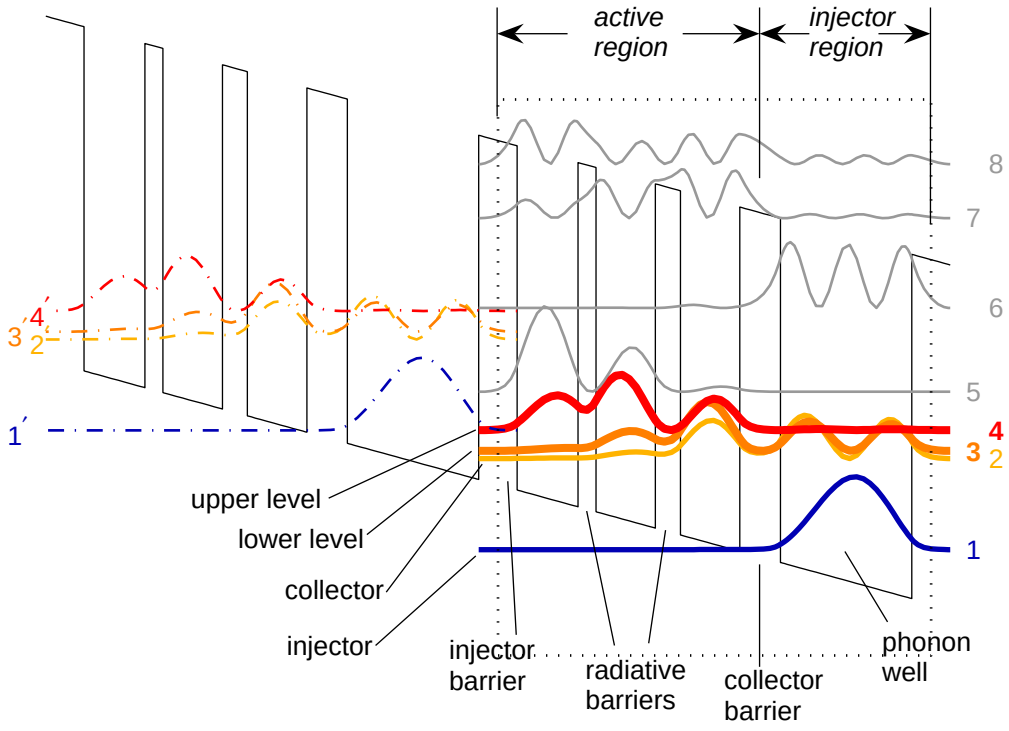


Figure 3-3: Sample band diagram of an OWI family device (design OWI185-M1 from [32, 53]); one QCL period is boxed. At the design bias, the injector $1'$ of the previous module aligns in energy with the upper laser level ($1' - 4$ anticrossing), populating the upper level through resonant tunneling. After photon emission ($E_{43} = 8.6\text{meV} \equiv 2.1\text{THz}$), the lower level is depopulated by resonant tunneling extraction by the collector, which is the excited state of an adjacent well. E_{21} is 35.8meV, approximately resonant with the GaAs LO phonon energy of $E_{LO} \approx 36\text{meV}$. This in turn enables depopulation of the collector via fast LO phonon scattering into the injector level 1.

multi-level injector with several closely spaced subbands can suffer from significant photon absorption. For example, in FL designs, this large injector population makes reabsorption of photons by the two energetically close double injector subbands a serious problem at low photon energies. Using only one injector level avoids this problem. This problem is even worse in CSL and BTC designs, since the miniband injectors they employ involved many subbands—although ironically, injector level absorption notwithstanding, BTC designs hold the record for low frequency operation [17].

Belkin et al. [54] and then Kumar et al. realized soon afterward that OWI designs could potentially be advantageous for high temperature performance as well. A single subband injector is energetically narrowest, resulting in improved injection selectivity. Also, the single injector well significantly simplifies the high energy subband structure compared to the two-well injectors of the FL family of devices. This eases avoidance of upper level parasitics.

Indeed, the current temperature record of 186K is held by an OWI device: OWI222G (see section 4.2.5) uses a one-well injector along with a moderately diagonal radiative transition to achieve its good temperature performance.

3.1.3 DSL devices

The acronym "DSL" stands for "Diagonal" radiative transition and "Superlattice" active region, but like the FL family, this acronym is no longer meaningful. This family represents the IP variant of phonon depopulation based designs, the key characteristic of which being the elimination of resonant tunneling collection. Instead, the upper laser level depopulates directly into the excited state of the adjacent phonon well through a diagonal radiative transition, and the lower laser level is in turn directly depopulated by LO phonon emission. In principle, this ought to hasten depopulation of the lower laser level. Simultaneously, this increases the upper laser level lifetime through increased diagonality (see section 3.2.1). The band diagram of a sample DSL design is shown in figure 3-4.

For reasons not well understood, this family of devices has never performed well.

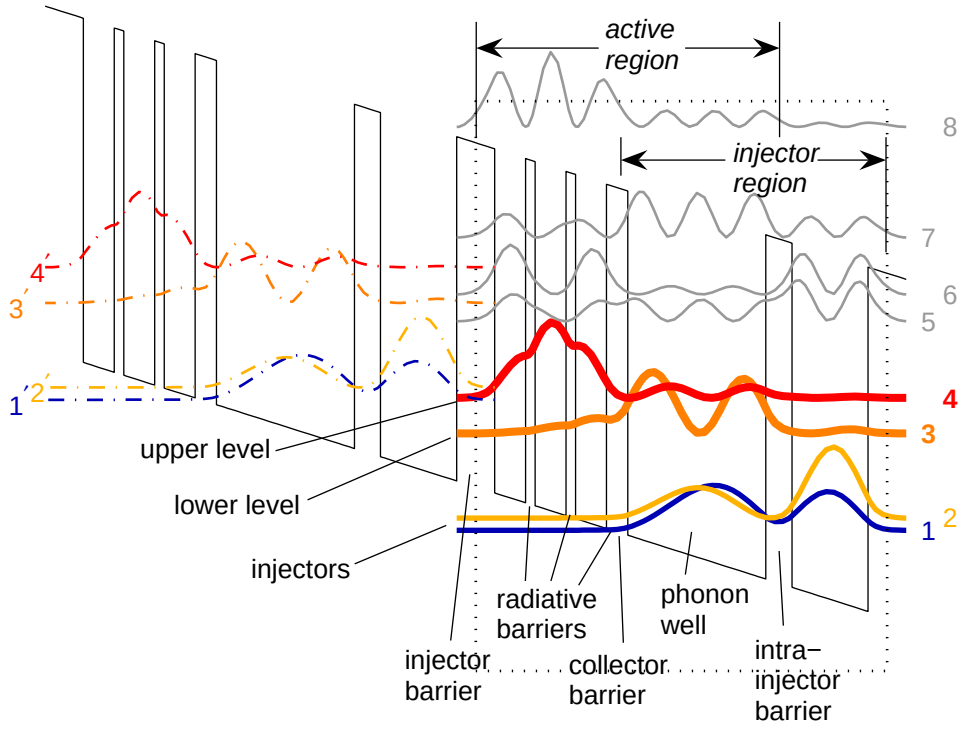


Figure 3-4: Sample band diagram of an DSL family device (design DSL203E-M1 from [32]); one QCL period is boxed. At the design bias, the injector $1'$ of the previous module aligns in energy with the upper laser level ($1' - 4$ anticrossing), populating the upper level through resonant tunneling. Unlike standard RP designs, the lower state is in the phonon well, so after photon emission ($E_{43} = 13.9\text{meV} \equiv 3.36\text{THz}$) the lower state is directly depopulated by LO phonon scattering into the injector levels 1 and 2. E_{32} is 33meV, approximately resonant with the GaAs LO phonon energy of $E_{LO} \approx 36\text{meV}$).

Notably, DSL designs seem prone to non-lasing wafers which have extremely large turn-on voltages before any conduction is observed.

3.2 Design Considerations

This section details design considerations for achieving high temperature lasing in THz QCLs. There are two requirements of design: first, lasing requires preserving sufficient gain at high temperatures to overcome cavity and material loses, and second, the superlattice must be electrically stable so that the design bias can be reached.

3.2.1 Suppression of non-radiative scattering: on the benefits of diagonality

As mentioned in the introduction, the postulated limit to high temperature performance is degradation of population inversion due to non-radiative scattering. Some of these are illustrated in figure 3-5. The most important non-radiative scatter-

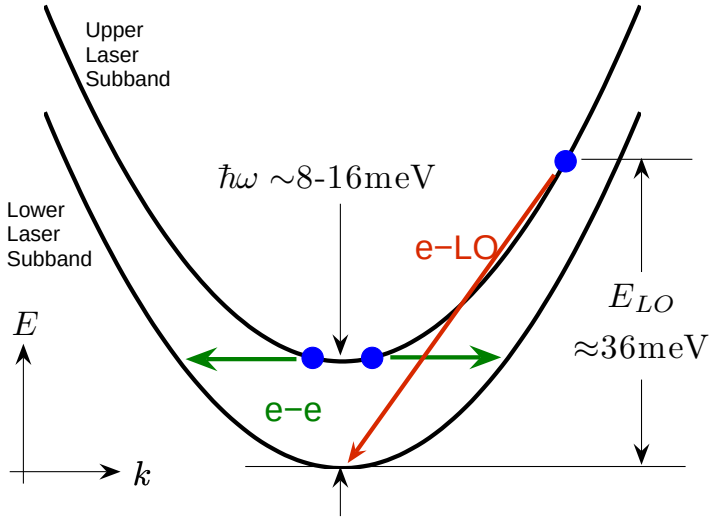


Figure 3-5: Non-radiative scattering mechanisms between upper and lower laser subbands. Electron-electron scattering is indicated in green. Thermally activated electron-LO-phonon scattering is indicated in red. k indicates 2D in-plane momentum.

ing mechanism is hypothesized to be thermally activated LO phonon scattering. A fundamental problem of THz QCLs is that radiative subband separations are necessarily below the semiconductor Reststrahlen band, the frequency range spanned by the transverse and longitudinal optical phonon frequencies ω_{TO} and ω_{LO}. Therefore, with increasing temperature, upper laser level electrons gain enough energy to transit to the lower laser level through LO phonon emission rather than the desired photon emission. The upper level lifetime decreases continually as temperature increases, so that ultimately insufficient population inversion remains for lasing. This problem is absent in mid-IR QCLs because the the radiative gap is much larger than the LO phonon energy, making scattering less temperature dependent.

That thermally activated LO phonon assisted non-radiative depopulation is a cause of laser degradation is experimentally supported by magnetic field assisted QCL operation experiments by Wade et al. [18]. Wade et al. used extremely strong magnetic fields ($\sim$16T) to produce in-plane confinement, hence quenching the in-plane free-electron like dispersion that enables thermally activated LO phonon scattering. This enabled them to reach lasing temperatures up to 225K. While such strong magnetic fields are impractical for real applications, this experiment suggests that higher operating temperatures might be achieved if alternative methods of protecting the upper level lifetime are found. To this effect, the MIT group has been pursuing the design of *highly diagonal QCLs* with ever decreasing oscillator strengths.

Recall from section 2.5.2 that peak optical gain associated with a given intersubband transition has the dependence

$$g_{\max} \propto \frac{\Delta N f_{if}}{\Delta \nu} \tag{3.1}$$

where ΔN is the population inversion, f_{if} is the oscillator strength, and $\Delta\nu$ is the gain linewidth broadening. Equation (2.48) shows that the LO phonon scattering rate is proportional to the scattering form factor given by equation (2.47), and equation (2.81) shows that the oscillator strength equation (2.81) is directly proportional to the dipole matrix element. Both the form factor and dipole matrix element are overlap

integrals of some sort between the upper and lower level wavefunctions. Unfortunately, this leads to the conflict that protecting population inversion from thermally activated LO phonon scattering demands that this overlap *decrease*, but increasing oscillator strength demands that it must *increase*.

Whereas historically THz QCL development actively designed for large oscillator strengths, the insight above motivates a search for an optimal balance in this trade-off between upper level lifetime and oscillator strength. In this, one may use the figure-of-merit defined in equation (2.84) as a guide. In FOM $= f_{fi}\tau_3\left(1-\frac{\tau_2}{\tau_{32}}\right)$, increasing diagonality (decreasing f_{fi}) lengthens τ_3 and τ_{32} and slows their decay with rising temperature. τ_2 is essentially set by the LO phonon depopulation gap, and is relatively constant with respect to temperature. Detailed calculations are deferred to chapter 4, which discusses actual THz QCL designs, but it turns out that the overall result is increased FOM at higher temperatures.

A major unknown, however, is the effects of gain linewidth, on which little information exists in general. Diagonal transitions are inherently more sensitive to interface roughness scattering, and so possess broader linewidths. Linewidths can also be temperature dependent; indeed, NEGF simulations by Nelander and Wacker suggest that temperature induced linewidth broadening may be limiting high temperature performance more than population inversion [55], although this interpretation remains highly speculative. However, one hope is that if linewidth in a diagonal device starts broader, relative change in linewidth versus temperature will be smaller compared to vertical designs. Therefore, a diagonal design could potentially be more robust to temperature induced linewidth broadening.

3.2.2 Thermal backfilling

Another concern at high temperatures is thermal backfilling from the injector level(s) to the lower laser level. At high temperatures, the QCL gain medium becomes sufficiently populated with LO phonons that reabsorption occurs between the collector and injector levels. In this case, instead of the upper laser level getting depopulated non-radiatively, the lower level becomes thermally populated, again reducing

population inversion. While this is a recognized problem in mid-IR lasers operating above room temperature, there is less consensus on whether thermal back-filling is significant in THz QCLs still operating below 200K.

This question inspired the design of several of lasers in this thesis that use *two phonon depopulation*, where the single phonon depopulation gap in standard RP and IP designs is replaced by two (not necessarily both resonant) phonon gaps. Similar two-phonon designs are commonly used in mid-IR lasers to inhibit thermal backfilling (although more often these are doubly resonant; for example, see [19]).

3.2.3 Negative differential resistance

Negative differential resistance (NDR) occurs whenever the small-signal resistance of a device is negative: at a given bias, a further increase in bias results in a drop in current. Studies of superlattice transport suggests that superlattices in NDR are generally inhomogeneously biased [28, 56], with different modules breaking into "domains" of differing electrical field strengths and possibly oscillating with time. This breaks the QCL electron recycling scheme, thus reducing gain.

In lasing QCLs, there is normally at least one major NDR feature, that which corresponds to the misalignment of the injector and the upper laser levels. When this misalignment occurs, lasing abruptly ends, and the current drops. But, as discussed in the next section, it is also possible for other NDR features to occur ahead of the threshold bias point. Lasing in NDR in this manner is believed to be detrimental to temperature performance.

3.2.4 Parasitic current channels

In a given QCL design, there is normally only one desired sequence of quantum states through which electron flow results in gain at a targeted radiative transition. All other paths of current conduction are undesirable and are termed *parasitic current channels*, for they draw current but produce no radiation. Parasitic channels may be

divided into *upper level parasitics* and *lower level parasitics*.[3] The former are current flows that bypass the designed channel by transport through a subband with energy higher than the upper laser level. The latter bypass through a subband with energy lower than the upper laser level.

While upper level parasitics can be avoided in some cases by design, lower level parasitics are generally unavoidable (although they can be strongly suppressed). The injector-to-collector lower parasitic resonance manifests in practically all QCLs, because the injector level starts at zero bias as the lowest energy level. Bias is increased to bring the injector into resonance with the upper laser level, and along the way the injector come into resonance with all energy levels below the upper laser level. When the injector is resonant with the collector, it is believed that a strong parasitic current channel opens up consisting of successive resonant tunneling and fast LO phonon scattering steps. This is illustrated in figure 3-6.

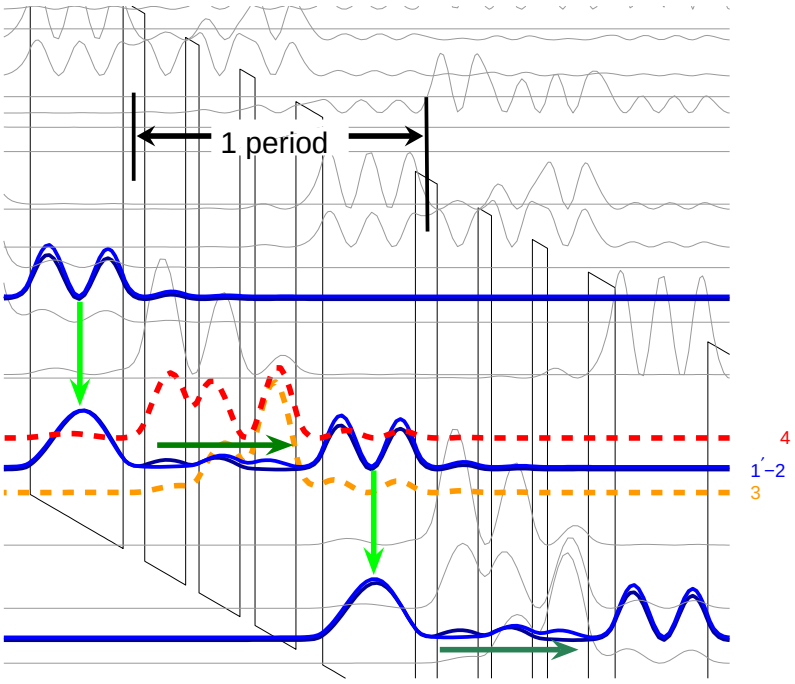


Figure 3-6: Schematic of the lower level parasitic due to injector-collector resonance ($1' - 2$ anticrossing) in an OWI design. The current path is indicated by the green arrows, with light green indicating LO phonon emission and dark green indicating resonant tunneling. The upper and lower laser levels 4 and 3 are entirely bypassed.

[3]Not to be confused with the upper and lower laser levels.

Parasitic lower level resonances divert current away from the radiative channel, and therefore increase the lasing threshold (a greater overall current is needed to fuel enough electron flow in the radiative channel). An extremely strong parasitic causes an early NDR as the lattice is biased beyond the parasitic resonance. A major goal of QCL design is keeping lower level parasitic currents low while still keeping peak currents high.

3.2.5 Injection and collection selectivity

Injection selectivity is the ability of injector levels to funnel electrons into the upper laser level without leakage into other states; in particular, funneling into the lower laser level should be avoided. *Collection efficiency* is the ability of the collector states to depopulate the lower laser level without drawing electrons from the upper laser level. Selectivity in both injection and collection is key to creating population inversion. The condition of *unity injection* is when all current flows through the radiative path and parasitics are avoided.

Unity injection is often assumed to keep transport calculations tractable, but is patently false in THz devices. Injection and extraction selectivity are particularly difficult for THz QCLs because typical radiative gaps ($\sim 5-15$meV) are of the same order as as the energy level broadening of the subband states. This makes it difficult to selectively direct transport into any one subband. Therefore, resonant tunneling injection demands keeping the injection anticrossing smaller than both the energetic broadening of the upper and lower laser levels and also the radiative gap. Yet a large anticrossing is desirable to maximize electron transport, and hence laser gain. In this trade-off, the injector barrier thickness producing the optimal injection anticrossing varies from design to design and is largely determined experimentally [57]. Similar concerns hold for the collector anticrossing, where a large anticrossing is desirable for fast depopulation but a small anticrossing is desirable for selective depopulation.

One notes that in principle, injection selectivity may be improved by by spatial separation of the upper and lower levels. This is another manner in which diagonality should improved performance.

THz QCLs have been historically limited to maximum operating temperatures of $\sim \hbar\omega/k_B$ [24], where $\hbar\omega$ is the photon energy. This trend which remains largely unexplained, but Kumar has suggested that it might be an artefact of resonant tunneling injection.[4] The argument is as follows: positive gain can only be attained when the injector levels are biased between the upper and lower radiative subbands, so low frequency designs have an inherently smaller usable bias range. Level broadening exacerbates the problem. Whether this explanation is satisfactory remains to be seen.

3.2.6 Doping

The QCLs in this thesis are grown in the $Al_xGa_{1-x}As$ material system and doped with silicon (Si), which is known to be an amphoteric dopant in GaAs. However, THz QCLs are typically very lightly doped (bulk concentrations of $\sim 5 \times 10^{15}\text{cm}^{-3}$), and Si doping is n-type at low doping concentrations. Typical designs are either bulk doped or δ-doped. Impurity scattering is sure to contribute to dephasing wherever dopants are located, so location is important. Typically, one avoids doping in the injector barrier or active region to avoid dephasing carrier injection or the radiative transition. Usually, doping is placed in one of the injector wells, but δ-doping is frequently placed in the collector barrier as well.

Dopants are assumed fully ionized, so that population inversion, and hence gain, is proportional to the doping density. However, increasing carrier density also increases free carrier losses and impurity scattering, which increases dephasing. The scaling of these detrimental effects is not so well understood. Empirical studies have confirmed that a given design has an optimal doping density [58], although good methods of predicting this value are unavailable. For the design studied in [58], doping was found to affect $T_{\max}$ by as a much as $\sim$25K over δ-doping concentrations ranging from $3.0 \times 10^{10}\text{cm}^{-2}$ to $5.0 \times 10^{10}\text{cm}^{-2}$, with the doping level of $3.6 \times 10^{10}\text{cm}^{-2}$ yielding the highest $T_{\max}$. The lasers of this thesis typically employ $3.0 \times 10^{10}\text{cm}^{-2}$ δ-doping or bulk equivalent.

[4] Unpublished results.

Chapter 4

Characterization, Measurement, and Experimental Results

4.1 Experimental Methods

4.1.1 Growth and fabrication

The MBE wafers of this study were grown by Dr. John Reno at Sandia National Laboratories. The accuracy to which wafers are grown is characterized using X-ray diffraction. This method probes the average periodicity of the superlattice rather than detailed layer by layer information. As such, growth inaccuracies are not necessarily integer monolayer multiples. A given design may be grown more than once, so each MBE wafer is uniquely numbered.

Wafers are processed into Fabry-Perot ridge lasers clad in copper-copper waveguides using methods described in [32]. Ridges were processed either using dry or wet etch methods, resulting in straight or sloped sidewalls, respectively. One notes that the sloped sidewalls of wet-etched devices introduces some ambiguity in the device width, and hence area, to use for the calculating current densities. The conventional adhered to in this thesis is to use the widest width, approximated by adding 20μm to the lithography mask metal width. Further fabrications details can be found in [32, 46].

4.1.2 Characterization

In the characterization of QCLs, there are three basic sets of experimental data. Current versus voltage is measured to form the IV. Optical power output versus current is measured to form the LI. Finally, for lasing devices, the *spectra* of the output radiation is measured to determine the lasing frequency. Typically, measurements are performed with pulsed mode biasing in order to avoid heating the device significantly above heat sink temperatures. Sometimes it is also useful to measure the differential conductance versus voltage or current to form the GV or GI. Loosely speaking, IV and GIV fall under "transport measurements," while LI and spectra fall under "optical measurements."

Ridge lasers were cleaved of length ∼1-2mm. Substrate backsides are gold-coated during fabrication, and this is used to indium solder the devices to custom made copper chip-carriers. Electrical connections are made using 25μm gold wire bonds. If transport data for a non-lasing device was desired, the device would be encapsulated in Stycast 2850 epoxy with LV23 catalyst following wire-bonding. Said chip carriers are then mounted on the cold stage of a pulsed-tubed cryocooler for measurement (Cryomech PT810). Devices intended to lase were fitted with Winston cones to improved light collection. Devices were biased either using a 2Ω/50Ω pulsed supply (Avtech AV-1011-B) or a DC supply (HP6632A). Typical pulsed biasing consisted of 300ns square pulses at a repetition rate dependent on the choice of optical detector.

The usual choice of optical detector was a helium-cooled Ge:Ga photoconductive detector (Infrared Laboratories unit #2189), for which 10kHz pulsing was usual. The Ge:Ga photodetector is based on valence band-to-impurity state excitations, limiting optical detection to above ∼2THz. For optical frequencies less than 2THz, either a helium-cooled InSb hot electron bolometer (MM-IR spectra, Inc.) or a helium-cooled Si bolometer was used (Infrared Laboratories), for which 3kHz and 300Hz pulsing was usual, respectively. The InSb bolometer has a high frequency roll-off at optical frequencies of ∼1.5THz. The Si bolometer is broad-band, with an optical low-pass filter cutting frequencies above ∼5THz. A thermopile was used for continuous wave

(CW) measurements (Coherent P4-42), with laser output chopped at typically 100Hz. All four detectors mentioned above were fitted with 1.9mm diameter Winston cones for light collection. Pulsed optical signal or chopped CW signal was measured by lock-in detection (EG&G 5209). Absolute power measurements were made using a calibrated thermopile detector (Scientech AC2500). One notes that power measurement for THz QCLs is in general fraught with uncertainties of collection efficiency, and no corrections are made for these.

Pulsed current was detected using a calibrated inductive current sensor, and pulsed voltage was detected using a resistive voltage divider. Each of these were sampled using a boxcar averager (Stanford Research Systems SR 250 or Princeton Applied Research Model 4121B). CW current was measured using a $\sim 0.5\Omega$ sense resistor, and CW voltage was measure directly across the device. While in principle one may obtain GIV data from numerical differentiation of the IV data, this method can be extremely noisy. To directly probe GV, a small sinusoidal bias (typically 100kHz) was applied on top of a CW bias using a bias-T (Pasternack PE1611), and a lock-in amplifier was used to detect the resulting small-signal current oscillations.

Spectra were measured using a Nicolet Magna 850 infrared spectrometer in linear scan, typically purged with dry nitrogen to limit atmospheric absorption. Either one of the three pulsed optical detectors mentioned above was used, or a deuterated triglycine sulfate (DTGS) detector internal to the spectrometer.

The data collection process is PC controlled via GPIB, with measurements collected by digital multimeters (HP34401A), or sometimes directly from the measurement instrument if it had a GPIB interface (eg. the lock-in amplifiers).

One contribution, albeit minor, of the present thesis was improving the data collection process. Prior to the author's arrival in the MIT THz group, one generally needed to collect IVs, LIs, and GIVs from separate measurements, and cyrocooler temperature was controlled manually. Spectra needed to be collected manually for different biases. This work introduced methods for pulsed "$LIVT$" and CW "$GLIVT$" measurements, where most quantities are collected simultaneously and the task of controlling the cyromech temperature is handled by the PC. Spectra acquisition was

also automated. These improvements were necessary for the collection of increasingly detailed transport information. In particular, previous THz QCL investigations rarely collected pulsed IVs, relying instead on CW IVs versus temperature. Pulsed IVs are now standard.

The next sections discuss in more detail some quantities of experimental interest: the *threshold current density* (J_{th}) in section 4.1.3, *maximum current density* (J_{max}) in section 4.1.4, *dynamic range* in section 4.1.5, and *differential conductance* (GIV) in section 4.1.6.

4.1.3 Threshold current density (J_{th})

J_{th} is the current density at which the QCL starts lasing. This manifest as a discontinuous increase in the current versus voltage slope when J_{th} is reached; the slope increases due to the increased transport caused by stimulated emission. This "kink" has varying degrees of visibility: it is obvious in devices with a prominent $1' - 2$ parasitic, but can be nearly invisible for devices with suppressed parasitics. The magnitude of the slope discontinuity also has a theoretical interpretation to be discussed in section 4.1.6. J_{th} is important because it is proportional to the population inversion required to achieve lasing (up to some temperature dependent prefactor). That said, it is a crude diagnostic; J_{th} in most cases rises with temperature, but it is difficult to pinpoint the physical mechanism responsible. Possible causes of reduction in gain include

- thermally induced population inversion degradation, in which case more pumping is required to achieve the same population inversion.
- thermally induced linewidth broadening, reducing peak gain.
- thermally induced increase of device loss (due to thermal redistribution of carriers between different quantum states in the gain medium).

In good designs, the rise of J_{th} versus temperature is usually characterized by a plateau region characteristically extending to $\sim 50-100$K, followed by a region of exponential

increase. Therefore, for sufficiently high temperatures, $J_{\rm th}$ is empirically well modeled by $J_{\rm th} \propto \exp(T/T_0)$, where T_0 is a fitting parameter. The fitted value of T_0 can be used to compare different devices and different designs. T_0 has a tendency to be higher in lower frequency devices. This is cited as experimental evidence supporting the hypothesis of phonon-assisted non-radiative scattering as a cause of population inversion degradation with temperature.

Nevertheless, T_0 must be interpreted with caution. Even devices from the same fabrication and wafer can exhibit significant differences in T_0. Moreover, the extracted value of T_0 depends somewhat on the number of high temperature threshold points used for fitting. It also has some dependence on device size, particularly in wet-etched devices. On a theoretical note, the mathematical form of this exponential fit is strikingly unphysical. The mathematics of statistical mechanics always produces terms like $\exp(E_0/k_BT)$ (where E_0 is some sort of characteristic energy), never $\exp(T/T_0)$. Therefore, the physical significance of T_0 is unclear.

Finally, rather than fitting the highest temperature points to the expression $J_{\rm th} = J_1 \exp(T/T_0)$, another common procedure is to the fit all threshold data to $J_{\rm th} = J_0 + J_1 \exp(T/T_0)$. T_0 from these different fits are significantly different and should not be directly compared.

4.1.4 Maximum current density ($J_{\rm max}$)

$J_{\rm max}$ is the current density at which the main NDR occurs. After this point, injector-upper laser level misalignment presumably results in a sudden loss of current flow. Therefore, at least at low temperatures, $J_{\rm max}$ is a measure of injection-to-active region transport and is essentially determined by the injector barrier thickness. The behavior of $J_{\rm max}$ with temperature gives an indication of the coherence of injection. For example, equation (2.14) shows that in the strong-coupling limit $J_{\rm max}$ ought to be essentially be determined by the upper level lifetime (either lasing or non-lasing). Conversely, equation (2.15) shows that in the weak-coupling limit $J_{\rm max}$ ought to be essentially determined by the coherence of transport; the anticrossing coupling, Ω, is temperature independent, and one expects somewhat different dependences for the

dephasing time $\tau_{||}$ and the upper level lifetime.

$J_{\max}$ often —but not always— corresponds to the point of maximum optical power.

4.1.5 Dynamic range

Clearly, one always desire J_{th} to be small and $J_{\max}$ to be large. The difference $J_{\max} - J_{\text{th}}$ is known as the *dynamic range*. One qualitative way to think about QCL laser operation is that the laser's dynamic range shrinks with temperature, mainly according to T_0. Lasing action stops when the dynamic range is entirely consumed. For a given T_0, mid-IR QCLs have a significant advantage over THz QCLs, because they possess a much larger dynamic range (the problem is even worse because T_0 is generally lower for THz devices, owing to the sub-Reststrahlen band operation).

4.1.6 Differential conductance

Measurements of small-signal conductance G versus voltage or current (GV or GI measurements) yield two sorts of useful information. First, regions of NDR manifest as roughness in the GV or GI. This enables discrimination of regions of slight NDR that may not be apparent from an IV measurement alone. Second, it gives the magnitude of the slope discontinuity, ΔG, in the IV at threshold. Note that in the literature it is standard to work with the differential resistance $R = 1/G$ and the corresponding discontinuity ΔR, but there is no fundamental difference in doing so.

Based on the rate equation approach of section 2.2.2, and employing the assumption that the upper laser level population varies as $dn_3/dJ \propto R$, Sirtori et al. [59] have argued that the slope discontinuity is given by

$$\frac{\Delta G}{G_2} = \frac{\Delta R}{R_1} = \frac{1 - \tau_2/\tau_{32}}{1 - \tau_2/\tau_{32} + \tau_2/\tau_3} \tag{4.1}$$

where G_2 (R_1) is the conductance (resistance) just above (below) threshold. Therefore, the fractional discontinuity is a measure of depopulation efficiency, and approaches unity as τ_2 approaches zero.

Based on a simplified 3-level density matrix approach in the vein of the Kazarinov-Suris formulation discussed in section 2.2.3, and without resorting to Sirtori's assumption, Kumar has derived that alternate interpretation

$$\frac{\Delta R}{R_1} = \frac{2\Delta n_{\text{th}}}{n_{\text{tot}}} \left(\frac{1 - \tau_{21}/2\tau_{31}}{1 + \tau_{21}/\tau_{31}} \right) \approx \frac{2\Delta n_{\text{th}}}{n_{\text{tot}}} \tag{4.2}$$

where Δn_{th} is the population inversion above threshold and n_{tot} is the total population [60]. This has a major qualitative difference from Sirtori's formulation in that the slope discontinuity becomes a measure of the population inversion at threshold instead of approaching unity for vanishing τ_{21}.

However, one notes that due to the small number of subbands and other simplifying assumptions, neither model is universally applicable to QCL analysis. In particular, both models become broken in the presence of a large lower level parasitic. The experimental results in this chapter present several such examples. In the worst case, the subthreshold IV becomes flat, meaning the subthreshold conductance G_1 is zero. This yields the pathological result that the fractional discontinuity truly is unity.

4.2 Experimental Results

This section presents the experimental results gathered over the course of this thesis. The designs here were largely designed by Dr. Sushil Kumar, and the measurements mostly done by the author.

The THz QCL community still awaits a quantitatively accurate simulation tool for design; in the absence of such a tool, QCL designs progress by incremental changes to existing designs guided by the simpler theories of chapter 2 and 3. As such, it is an inherently sequential process. Therefore, this section groups designs into chronological "generations," as this best reflects the thinking process that went into each design. The "first" generation represents the designs that were under investigation at the time the author joined the Hu group at MIT, and each subsequent generation describes

the evolutions since then.

4.2.1 On the calculation of band diagrams and anticrossings

Each design is accompanied by a single-module *band diagram* and an *anticrossing diagram*. The former is used for lasing frequency, oscillator strength, and scattering rate calculations while the former is used to identify the anticrossing gaps discussed in section 2.2.3.

The band diagram for each design shows the approximate Wannier-Stark eigenstate envelope functions calculated by diagonalizing the Hamiltonian of the biased superlattice potential of a single QCL module plus the injector barrier of the next module, employing Dirichlet boundary conditions ($F(z) = 0$ at simulation boundaries). The modulus squared of the eigenfunctions are plotted offset by their eigenenergies. By convention, the superlattice potential tilts downwards from left to right. Normally, the design bias is chosen to be the bias at which the injector level of one module is optimally aligned with the upper laser level of the next module, as determined by the anticrossing calculations to be discussed below. Eigenstates at this bias are numbered from lowest to highest energy starting with the injectors; this numbering is retained even if the eigenstates cross over in energy at different biases. The upper and lower laser levels are writ in bold to distinguish them from other eigenstates.

Each band diagram reports the layer width sequence in monolayers; all lasers in this thesis are grown in the $Al_xGa_{1-x}As$ system, so 1ML≈2.825Å. The band diagrams presented below are not based on the intended designs, but correct for growth inaccuracies. For example, if X-ray diffraction data determines 2.0% overgrowth, the band diagram is calculated with all layer widths multiplied by 1.02. In assuming homogeneous growth inaccuracy, it should be understood that these band diagrams are true only in an "average" sense. Layers that are fractions of monolayers in width are of course physically impossible, but a detailed layer-by-layer account of growth inaccuracies is not only unavailable, but would render the calculation intractable even if it were.

Finally, the LO phonon scattering lifetimes reported in each band diagram are calculated assuming a lattice temperature of 4K and an electronic temperature of 100K. Two lifetimes are reported: $\tau_{\mathrm{LO},ul}$ considers scattering only from the upper laser level to the lower laser level, while $\tau_{\mathrm{LO},u}$ includes scattering to all levels below the upper laser level.

The anticrossing diagram plots energy differences between eigenstates in a two-module simulation as function of bias. The minima of these differences yield the anticrossing gaps which approximate the interaction potential between localized well states. A "prime" designation indicates a level from a previous module; for example, if the upper laser level is 4 and the injector level is 1, then the $1' - 4$ anticrossing designates the bias at which the injector level of the previous module is optimally aligned with the upper laser level of the current module, and its anticrossing gap is designed $\Delta_{1'-4}$.

4.2.2 First generation designs

These devices were measured earliest in this course of research; the transport data is not as complete as measurements for later devices.

OWI202D-M1, OWI180E

OWI202D-M1 is modification of the original OWI design (OWI185-M1, see reference [53]) changed to lase at a higher frequency. Device specifications and experimental data are presented in figures 4-1 and 4-2. Originally, it was postulated that the mechanism limiting temperature performance is population inversion degradation due to level 5 acting as an upper level parasitic. Level 5 is relatively close in energy to the upper level 4 and has a large overlap with the upper level envelope function, thus can be populated by LO phonon absorption.

To test this hypothesis, OWI180E was designed. In this modification of the OWI scheme, the active region well structure is modified so that former parasitic level 5 becomes the new upper level, and the former upper level 4 is brought into resonance

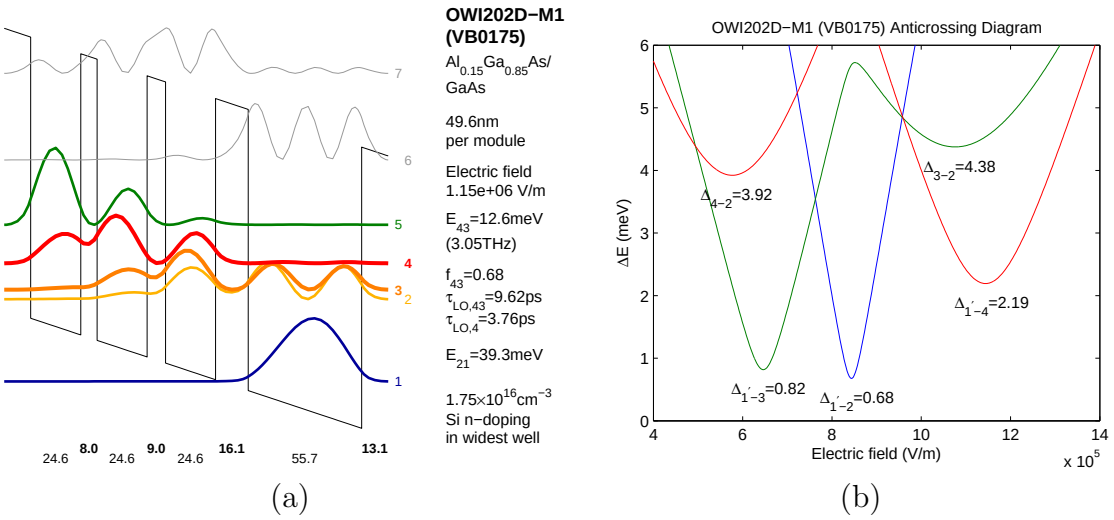


Figure 4-1: Design parameters for OWI202D-M1, wafer VB0175 (overgrown by 0.39%). (a) Calculated single-module band diagram at optimal injection bias. Subband 5 was postulated to degrade temperature performance by acting as an upper level parasitic for upper laser level 4. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}, and injection anticrossing is $\Delta_{1'-4}$.

with the lower level and collector. Design specifications and experimental data are presented in figures 4-3 and 4-4 respectively. With this subband structure, the next highest eigenstate becomes spatially isolated from the upper laser level, thus discouraging LO phonon absorption. Indeed, OWI180E did achieve superior performance, with $T_{\max} = 174$K as opposed to $T_{\max} = 155$K for OWI202D-M1.

This comparative study was thought to have confirmed the upper level parasitic hypothesis, but in retrospect, this conclusion may have been too hastily made. There are at least three problems hindering a definite conclusion. First, the phonon depopulation gap (E_{21}) was slightly larger in OWI180E (41.2meV; although later evidence suggests that thermal backfilling is not currently a performance barrier). Second, OWI180E was designed with a larger radiative gap than OWI202D-M1. The ratio of lasing frequencies between the two devices is roughly the same as the ratio of their $T_{\max}$, in accordance with the empirical $\sim \hbar\omega/k_B$ temperature limit. Third, T_0 was higher for OWI202D-M1 than it was for OWI180E (162K versus 96K), suggesting that OWI202D-M1 is actually *less* sensitive to temperature dependent effects; although

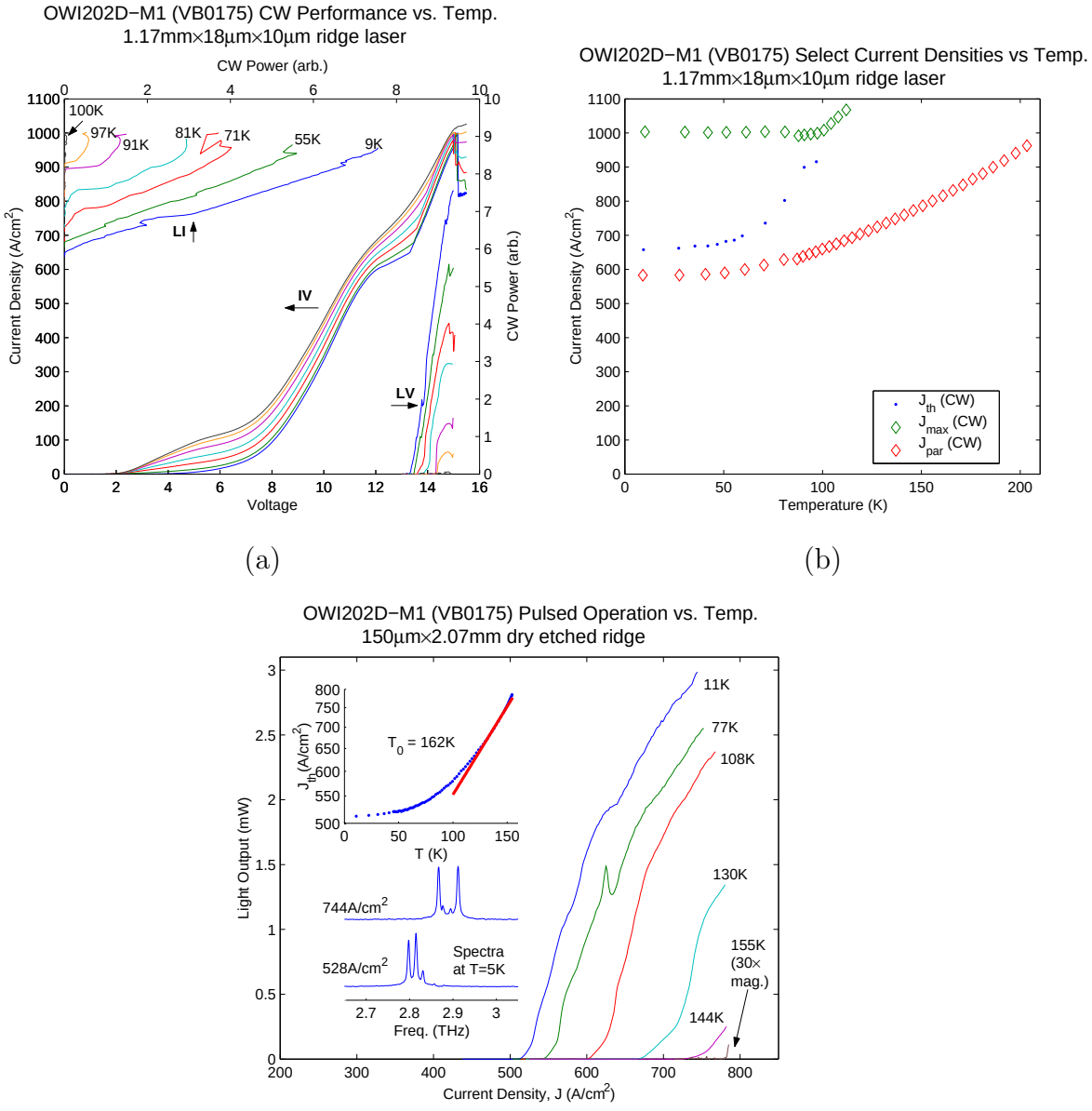
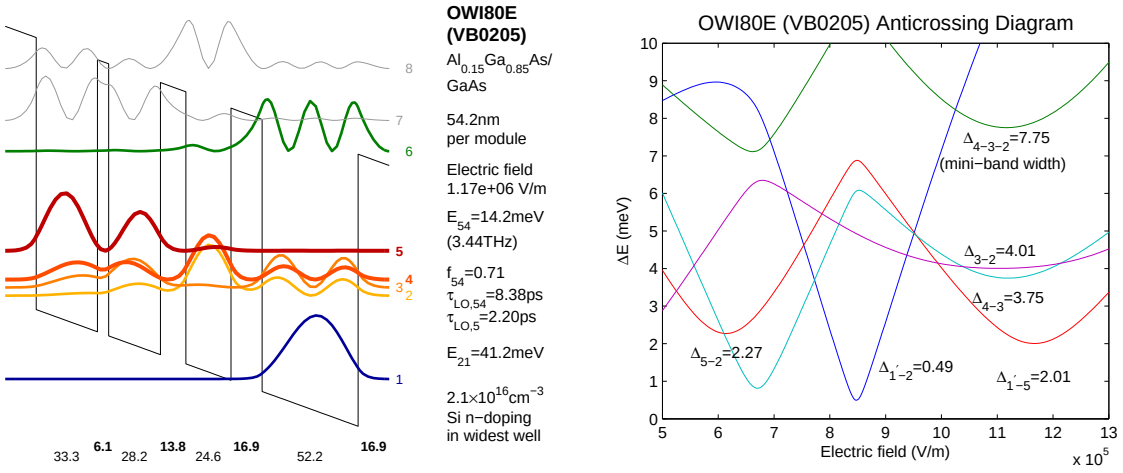


Figure 4-2: Experimental results for OWI202D-M1, wafer VB0175 (overgrown by 0.39%). (a) CW LIV vs. temperature measured for a narrow device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated at 11.7V applied bias. (c) Pulsed LI measured for a wide device lasing up to 155K.



(a) (b)

Figure 4-3: Design parameters for OWI180E, wafer VB0205 (overgrown by 2.4%). (a) Calculated single-module band diagram at optimal injection bias. The next highest energy level is both spatially and energetically isolated from the upper laser level, thus becomes unlikely to act as a parasitic. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector miniband width is Δ_{4-2}, and injection anticrossing is $\Delta_{1'-5}$.

this may be unsurprising due to its lower lasing frequency, and a slight departure from T_0=162K is noted at the highest temperatures.

An unusual feature of OWI180E is the current at its parasitic knee. While a prominent knee such as that seen in figure 4-4a is a common feature of many designs both before and after OWI180E, OWI180E has a parasitic knee that is amazingly independent of temperature (more so than other devices), beginning to rise only around 200K. One possibility considered was that the J_{par} versus temperature curve reflects the LO phonon population, to which 2-1 LO phonon scattering should be proportional. However, a fit to this curve using a Bose-Einstein distribution with an offset and scaling factor fits poorly. OWI180E also lased at significantly lower frequency than expected.

Summary

Whether the absence of an upper level parasitic is the reason for the improvement of OWI180E over OWI202D-M1 remains uncertain.

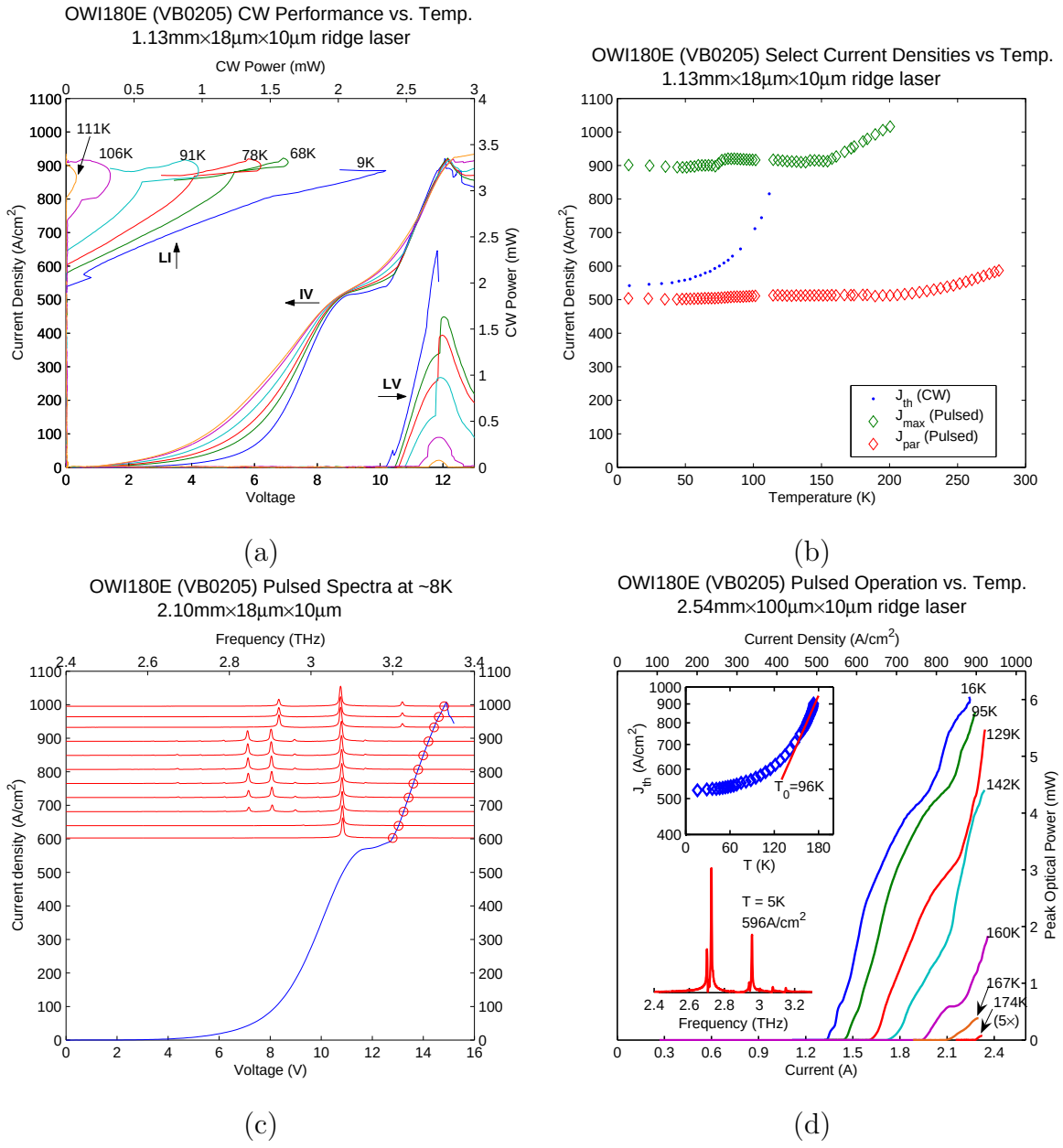


Figure 4-4: Experimental results for OWI180E, wafer VB0205 (overgrown by 2.4%). (a) CW LIV versus temperature measured for a narrow device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. Note that J_{th} is from a CW measurement, while J_{max} and J_{par} are from pulsed measurements; pulsed light data were not measured for the device in a. J_{par} is evaluated at 8.9V in the pulsed IVs. (c) CW spectra from a narrow device, overlaid on ∼8K IV to indicate bias location. Note that the spectra were measured for a device different than the one used in plots a and b, and the IV characteristics are hence somewhat different. (d) Pulsed LI and spectra data for a 174K device.

The improved performance of OWI180E over OWI202D-M1 (and also OWI185E-M1 in the next generation) also engendered speculation that thermal backfilling of the lower level is a potential source of temperature degradation. Evidence from third generation devices (discussed in section 4.2.4) suggests this is not the case, so it may be worth while to redesign OWI180E with a smaller phonon gap and all other parameters kept the same.

OWI202D-M1 is perhaps also worthy of further investigation, on account of its exceptionally high T_0.

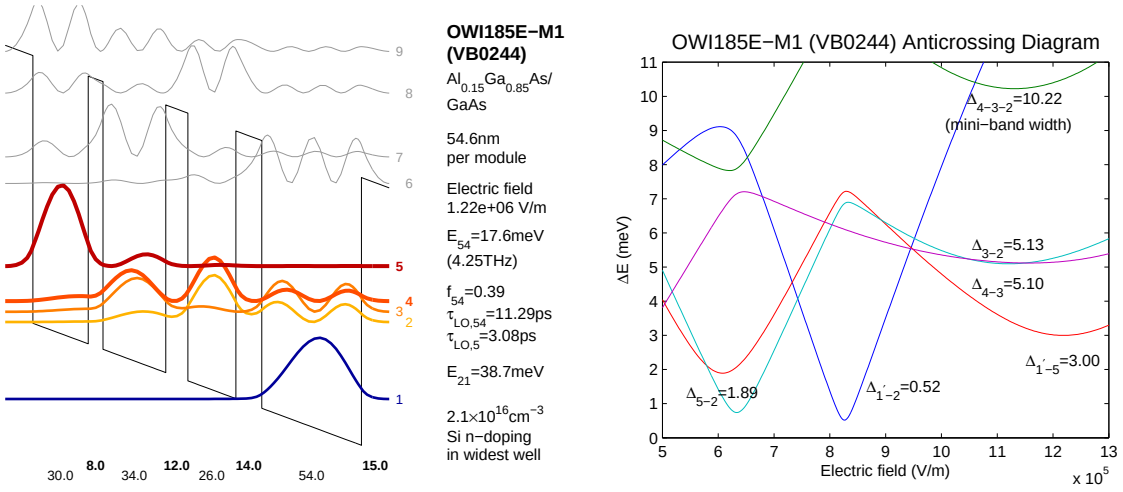
4.2.3 Second generation designs

This generation was the first to seriously explore the effects of increasing diagonality in the design. This second generation of devices has yielded three novel results: the first demonstration of LO phonon scattering assisted injection in a THz device, the highest temperature lasing to date, and the simplest QCL design to date.

OWI185E-M1

OWI185E-M1 is a modification of OWI180E, and one of the first forays into exploring the effects of diagonality. Design parameters are presented in figure 4-8. Between OWI185E-M1 and OWI180E, the oscillator strength is reduced from 0.74 to 0.39, the radiative gap is raised from 14.2meV to 17.6meV (3.44THz to 4.25THz), and both collector and injector anticrossings are increased. The phonon depopulation gap is also brought closer to LO phonon resonance, from 41.2meV in OWI180E to 38.7meV. So in short, *every* design parameter has been changed between OWI180E and OWI185E-M1—not such a good idea in retrospect, as this significantly complicates analysis.

The intended 5-4 transition performed reasonably well ($T_{\max} \sim 154K$), but did poorer than OWI180E. However, this design yielded unexpected low frequency lasing (∼1.8THz) at very high biases. This serendipitous lasing actually lased up to 163K, even higher than the originally designed transition. One hypothesis is that



(a) (b)

Figure 4-5: Design parameters for OWI185E-M1, wafer VB0244 (overgrown by 0.09%), for resonant tunneling injection. (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector miniband width is Δ_{4-3-2}, and injection anticrossing is $\Delta_{1'-5}$.

the lasing is due to Stark-effect splitting of the collector levels, with lasing occurring between subbands 4 and 3. Based on its lasing frequency, the postulated band diagram is shown in figure 4-6. Under this interpretation, the upper laser level is not resonant with any preceding energy levels. Therefore the high-bias upper laser level is hypothesized to be pumped by LO phonon *scattering assisted* (SA) injection. Scattering assisted injection has been proposed and implemented in mid-IR QCLs (see, for example, [61]), but OWI185E-M1 marks its first deployment in an THz device.

There are two SA possibilities: using the level numbering in figure 4-5, Kumar originally proposed that level 5, the upper laser level of the designed resonant-phonon lasing, becomes an electron reservoir for level 4, the upper laser level in the SA based lasing. In the course of writing this thesis, it was discovered that level 7 is highly anticrossed with level 3 of the previous module (anticrossing gap of 2.12meV), and 7-4 pumping may also be possible. Both of these are illustrated in figure 4-6. Figure 4-7 compares LO phonon scattering rates for these two injection mechanisms. The low temperature selectivity of the 5-4 injection mechanism is so poor that it becomes surprising that lasing occurs at all. The calculated 7-4 scattering rate is significantly

larger the 5-4 scattering rate at low temperatures, and is also much more selective. The disadvantage of the 7-4 scattering mechanism, however, is that the scattering rate is not expected to increase significantly with temperature, because of the large 7-4 energy separation. In contrast, figure 4-7 shows that injection selectivity increases with temperature for the 5-4 transition.

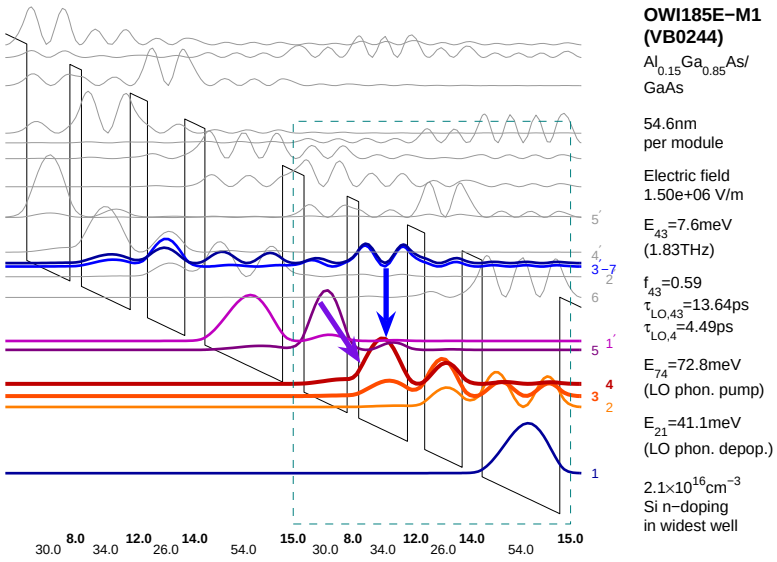
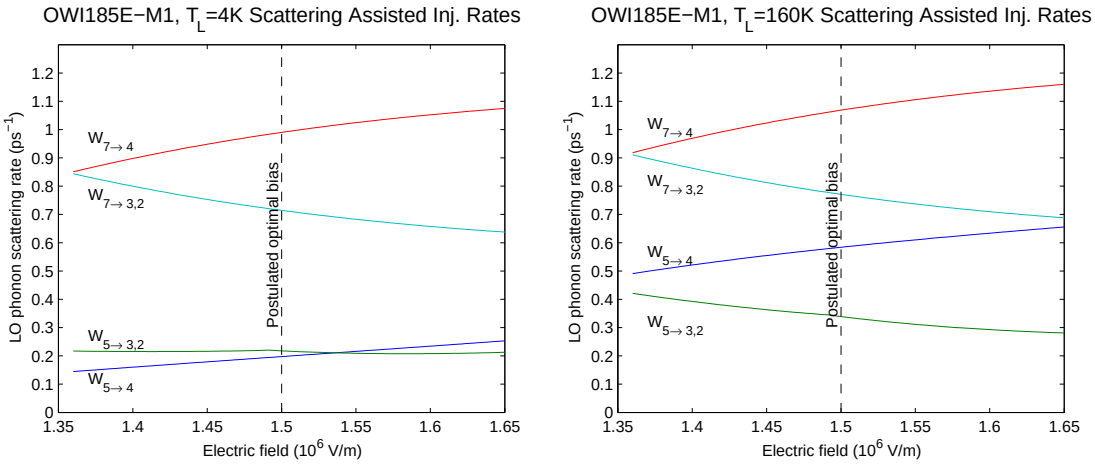


Figure 4-6: Postulated 2-module band diagram for scattering assisted lasing in OWI185E-M1, illustrating the possible 5→4 injection mechanism (purple) and the possible 7→4 injection mechanism (blue). The calculated $3' - 7$ anticrossing gap is 2.12meV.

Experimental data for both the resonant tunneling and scattering assisted transitions are presented in figure 4-8. Although the resonant tunneling and scattering assisted transitions are almost mutually exclusive, there exists a small bias region close to the main NDR where both frequencies lase simultaneously. Note, that the optical power of the 5-4 lasing peaks prior to the main NDR. This is consistent with a population build-up in level 4, such that both transitions start to lase prior to the onset of the main NDR.

Non-lasing current transport was also investigated for this design. As shown in figure 4-8d, the principle NDR disappears in the absence of lasing. Hence, one speculates that the NDR in the lasing device is related to the cessation of 5-4 lasing. Surprisingly, figure 4-8d also shows that the non-lasing IV has noticeably higher



(a) (b)

Figure 4-7: LO phonon scattering rates in OWI185E-M1 vs. electrical bias for 7-4 injection and 5-4 injection at (a) 4K (b) 160K. Rates are calculated based on a 1-module band diagram simulation. Electronic temperature is assumed to be 100K above the lattice temperature.

current in the vicinity of the parasitic knee. This is puzzling, as conventional wisdom holds that stimulated emission is irrelevant to current transport below threshold. The reason for the near-threshold differences seen in 4-8d are currently unknown, and has not been observed in the non-lasing measurements of other designs.

OWI222G

OWI222G currently possesses the highest lasing temperature of any published THz QCL gain medium. It has $T_{\max}$=186K, and a remarkably high T_0 = 158K given that it lased at the somewhat high frequency of 3.9THz. Design specifications and experimental data are presented in figures 4-10 and 4-11 respectively.

OWI222G has no predecessor amongst MIT designs, but Luo et al. published details of 3-level one-well injector design in 2007 ahead of OWI222G. Its design specifications, as calculated from layer widths reported in [62], are shown in figure 4-9. It lased up to 142K (Belkin et al. later used the exact same gain medium design and better waveguide fabrication to achieve 178K lasing [54]). In their design, Luo et al. adhered to the once conventional philosophy of making oscillator strengths as large as possible.

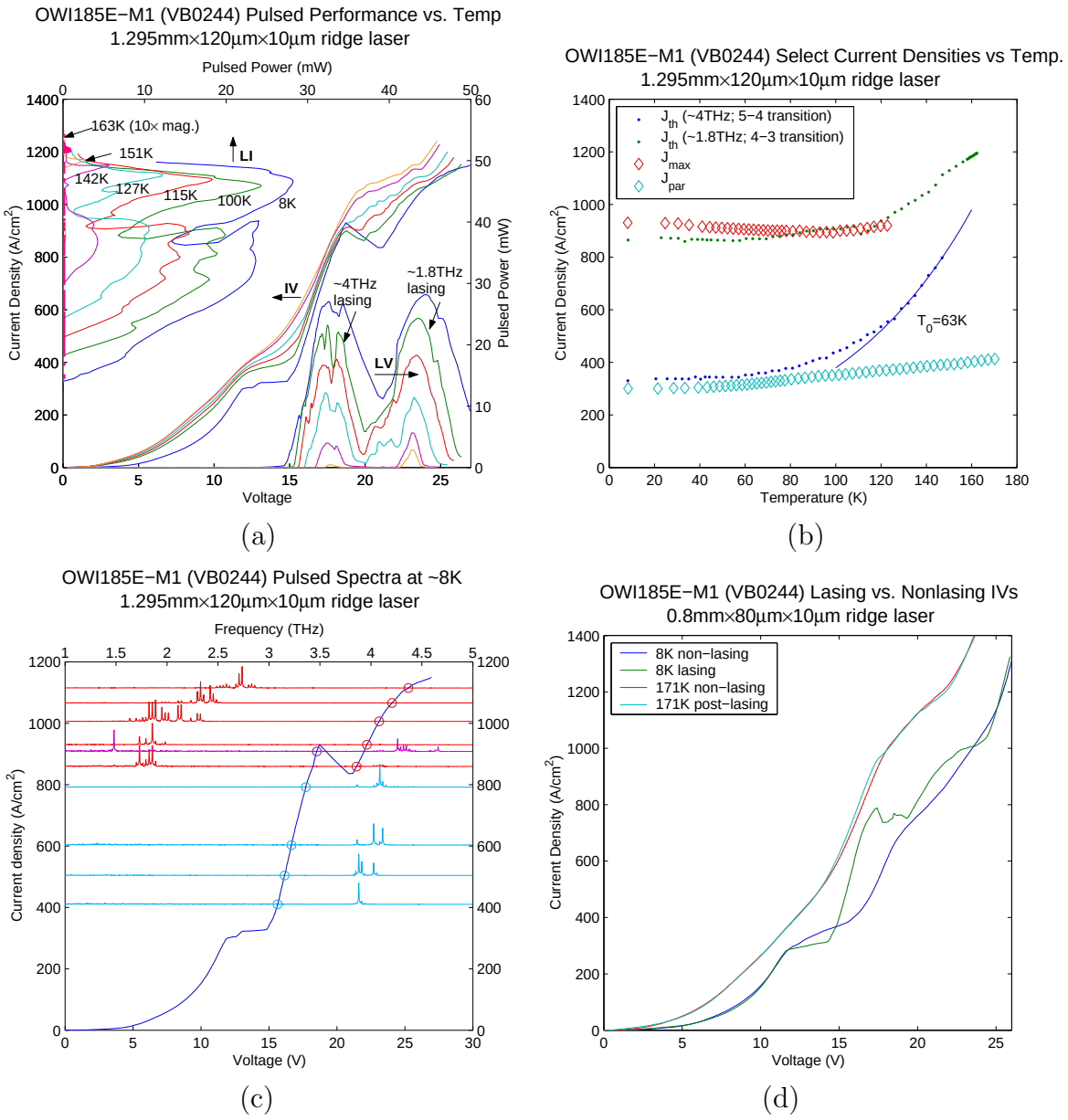
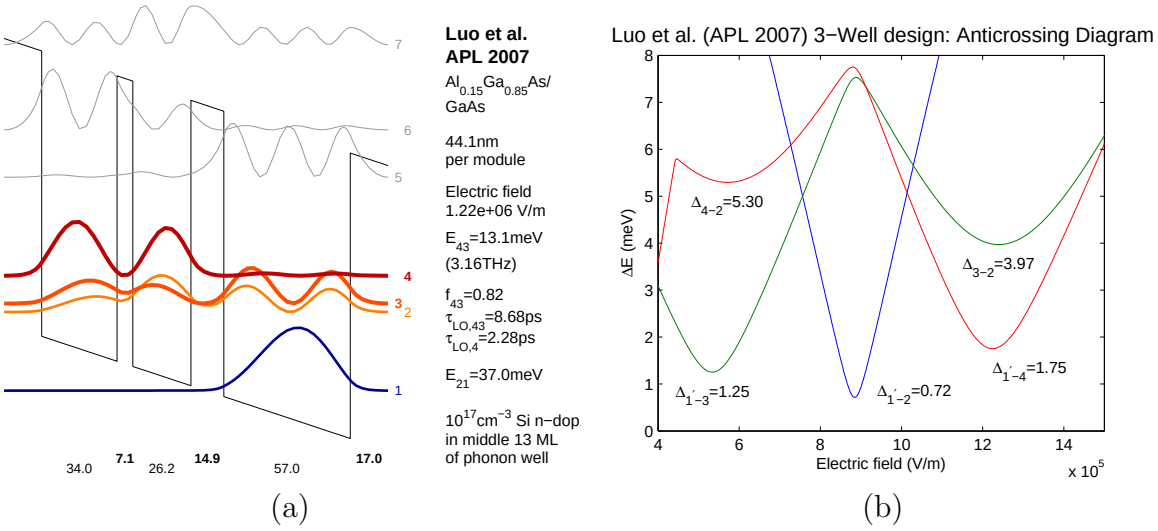
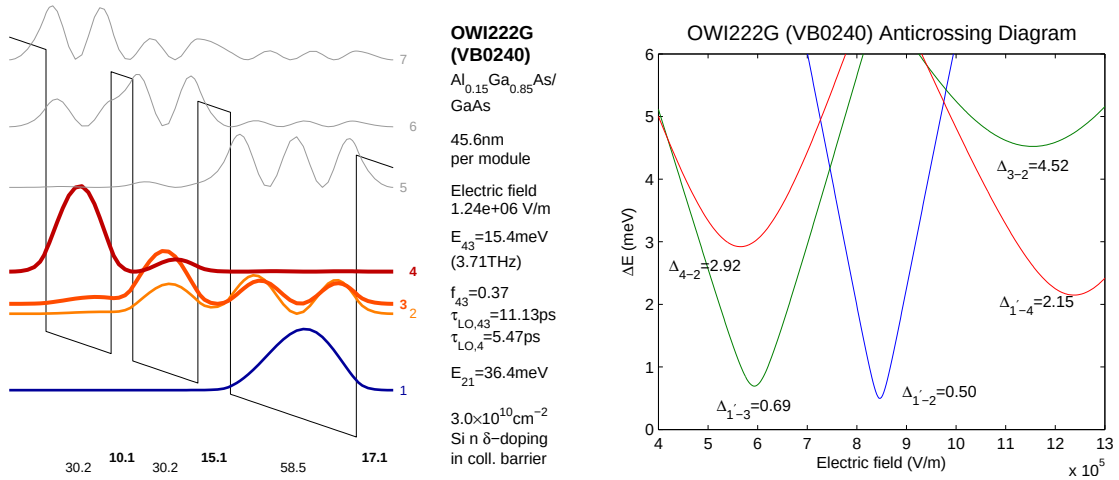


Figure 4-8: Experimental results for OWI185E-M1, wafer VB0244 (overgrown by 0.09%). (a) Pulsed LIV versus temperature measured for a wide device. Note that LV at 163K is missing because LI data was taken in a separate experiment from LV and IV. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. The threshold for the ∼1.8THz transition is determined by measuring the light output using a InSb hot electron bolometer. J_{par} is evaluated from the IVs at 12.0V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicated bias location. As indicated by the purple spectrum, close to NDR, both low and high frequency lasing occur simultaneously. (d) Lasing versus Non-lasing IV measurements. At high temperatures, IVs converge.



(a) (b)

Figure 4-9: Design parameters for design in reference [62]. (a) Calculated single-module band diagram at optimal injection bias. The oscillator strength is much higher than the value reported by Luo et al. ($f_{ul} = 0.51$) because they calculated oscillator strength for a two module simulation. (b) Calculated anticrossings. Collector anticrossing is Δ_{3-2}, and injection anticrossing is $\Delta_{1'-4}$.



(a) (b)

Figure 4-10: Design parameters for OWI222G, wafer VB0240 (overgrown by 0.79%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}, and injection anticrossing is $\Delta_{1'-4}$.

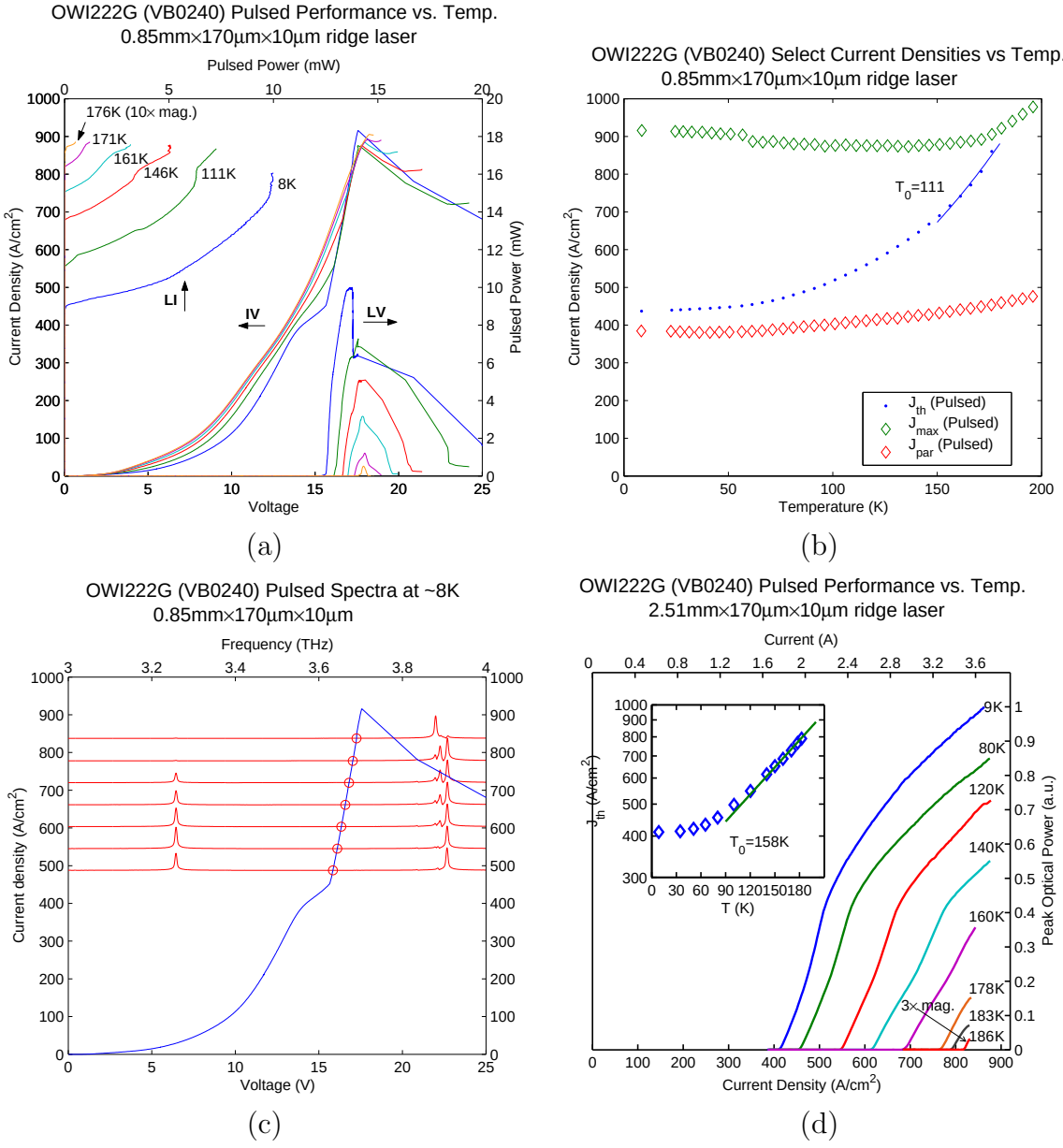


Figure 4-11: Experimental results for OWI222G, wafer VB0240 (overgrown by 0.79%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 13.9V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) Pulsed LI data for a wide device lasing up to 186K; no spectra available.

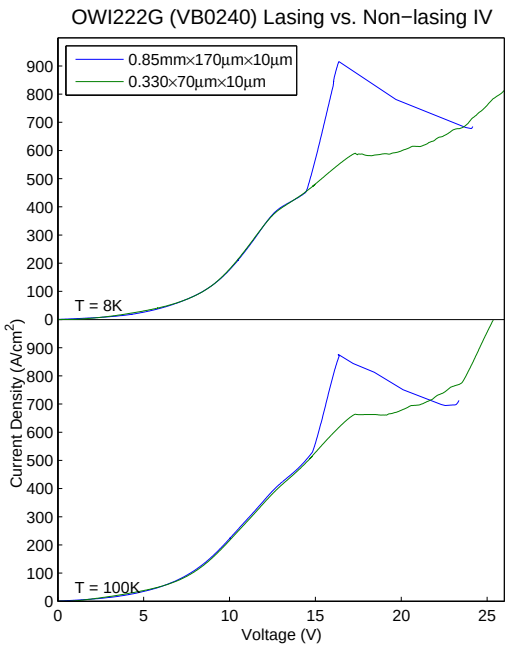


Figure 4-12: Lasing versus non-lasing IV measurements at ∼8K and ∼100K. Note that the lasing and non-lasing data are taken from different devices; a -1.2V shift has been applied to the lasing data, which is the same device from figures 4-11(a-c).

A comparison between the Luo et al. 2007 design and OWI222G is useful for demonstrating the theoretical benefits of diagonality. Figure 4-13 compares the figure-of-merit defined in equation (2.84) for the two designs. The upper level lifetime is taken to be

$$\frac{1}{\tau_u} = \frac{1}{\tau_{43}} + \frac{1}{\tau_{42}} + \frac{1}{\tau_{41}} \tag{4.3}$$

and lower level lifetime to be

$$\frac{1}{\tau_l} = \frac{1}{\tau_{31}} + \frac{1}{\tau_{21}} \tag{4.4}$$

and the upper-to-lower level lifetime is

$$\frac{1}{\tau_{ul}} = \frac{1}{\tau_{43}} + \frac{1}{\tau_{42}} \tag{4.5}$$

So the figure-of-merit of equation (2.84) becomes

$$\text{FOM} = f_{43}\tau_u \left(1 - \frac{\tau_{ul}}{\tau_l}\right). \tag{4.6}$$

Although the large reduction of oscillator strength from 0.82 to 0.37 yields only a modest improvement to the FOM, figure 4-13 shows that this improvement persists across all temperatures. For equal doping levels, device losses, and linewidths, the FOM becomes a direct measure of gain; figure 4-13 also indicates an extrapolated $T_{\max}$=211K for OWI222G based on these assumptions. Because the FOM tapers more slowly at high temperatures, even modest improvements should, in principle, enable significantly higher $T_{\max}$.

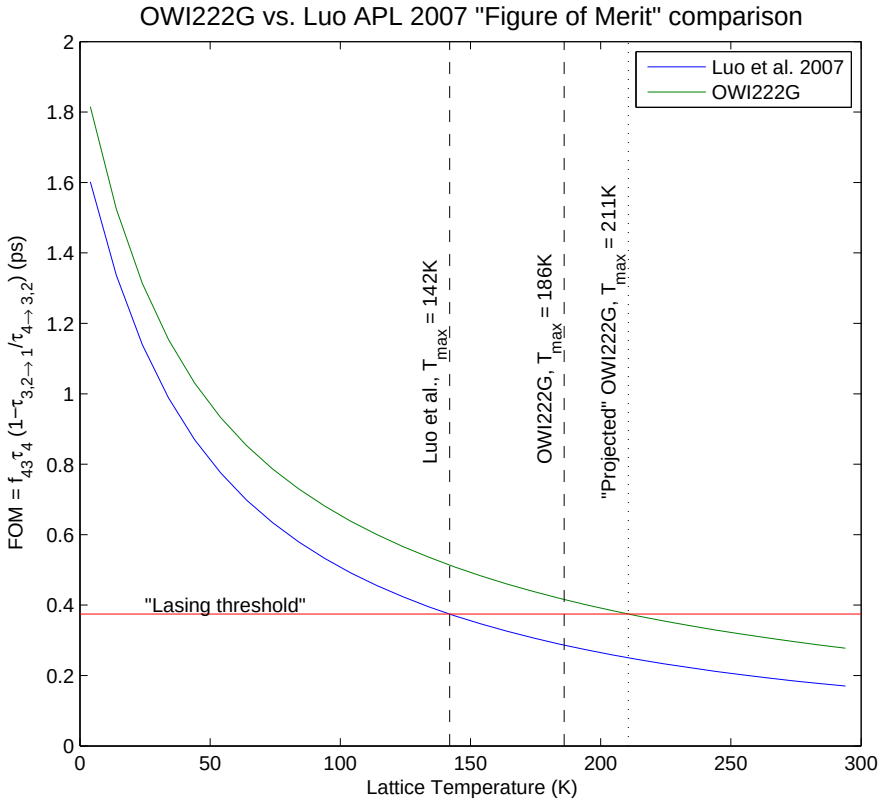


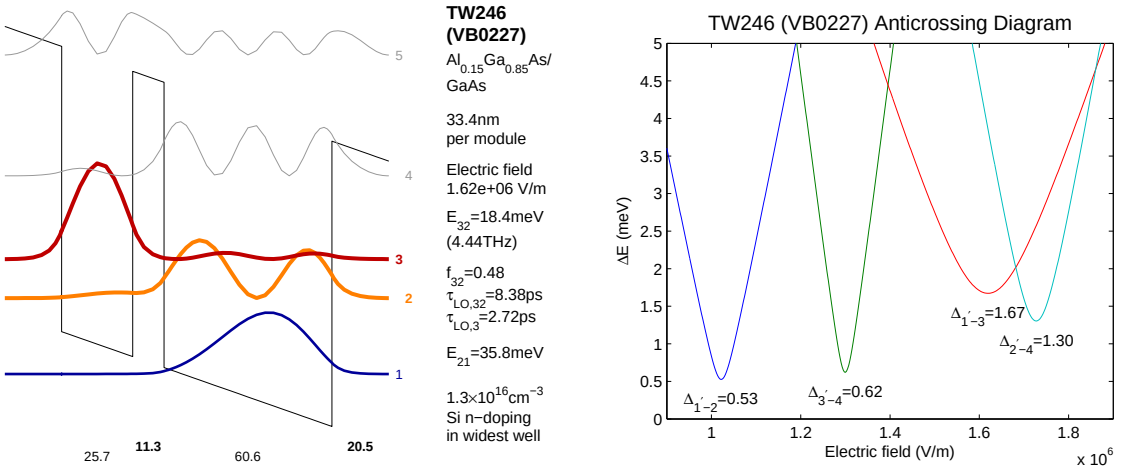
Figure 4-13: Comparison of the figure-of-merit defined in equation (2.84) between OWI222G and the Luo et al. APL 2007 3-well design. Only LO phonon scattering is considered, using the thermally-activated approximation of equation (2.54). The electronic temperature used in these calculations is assumed to be 100K above the lattice temperature. The pseudo-"lasing threshold" is determined from the intersection of the blue curve and the experimental $T_{\max}$ of Luo et al.

That said, the above analysis is of only qualitative importance, as the assumptions of equal doping levels, device losses, and linewidths are almost certainly false. For example, in figure 4-13, dividing the green line by ∼1.1 (ie. 10% broader gain

linewidth than the Luo et al. design) is enough to bring it into intersection with the red line at the experimentally observed 186K. And one must again bear in mind that Belkin et al. achieved $T_{\max}$=178K using the exact same gain medium design.

TW246

Although not labeled as such, TW246 is a DSL family device. "TW" in this case stands for "two-well." This two-well QCL is the simplest QCL in existence, and could possibly be the simplest QCL possible. A one-well QCL, as originally proposed by Kazarinov and Suris, has never been demonstrated due to NDR induced electrical instability. Design parameters and experimental results are shown in figures 4-14 and 4-15, and are further published in [63].



(a) (b)

Figure 4-14: Design parameters for TW246, wafer VB0227 (overgrown by 2.74%). Note that, in contrast to [63], the band diagram here is calculated without non-parabolicity. (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Injection anticrossing is $\Delta_{1'-3}$. As this is a DSL design, there is no collection anticrossing.

TW246 and OWI222G may be compared using the same figure-of-merit analysis used in the previous section to compare the design of Luo et al. 2007 to OWI222G. The results are presented in figure 4-16, according to which TW246 should have lased up to room temperature (the average bulk doping is the same in the two designs). Instead, TW246 performed significantly worse, lasing only up to $T_{\max}$=121K.

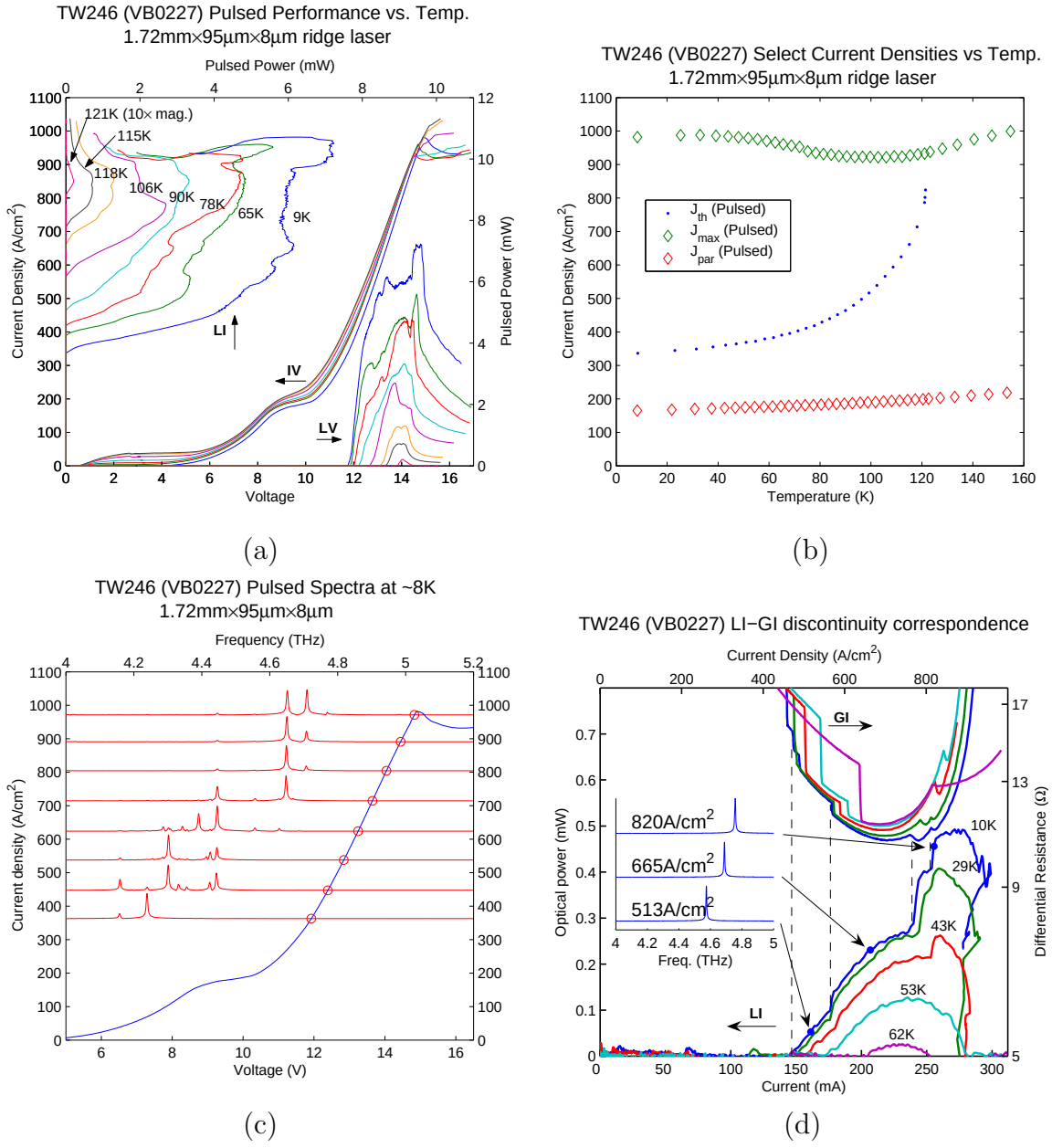


Figure 4-15: Experimental results for TW246, wafer VB0227 (overgrown by 2.74%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 8.9V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) Comparison of LI versus GI. Discontinuities in differential resistance match up with features in the LI; mode hopping is the apparent cause of both, as indicated by the spectral changes (see inset).

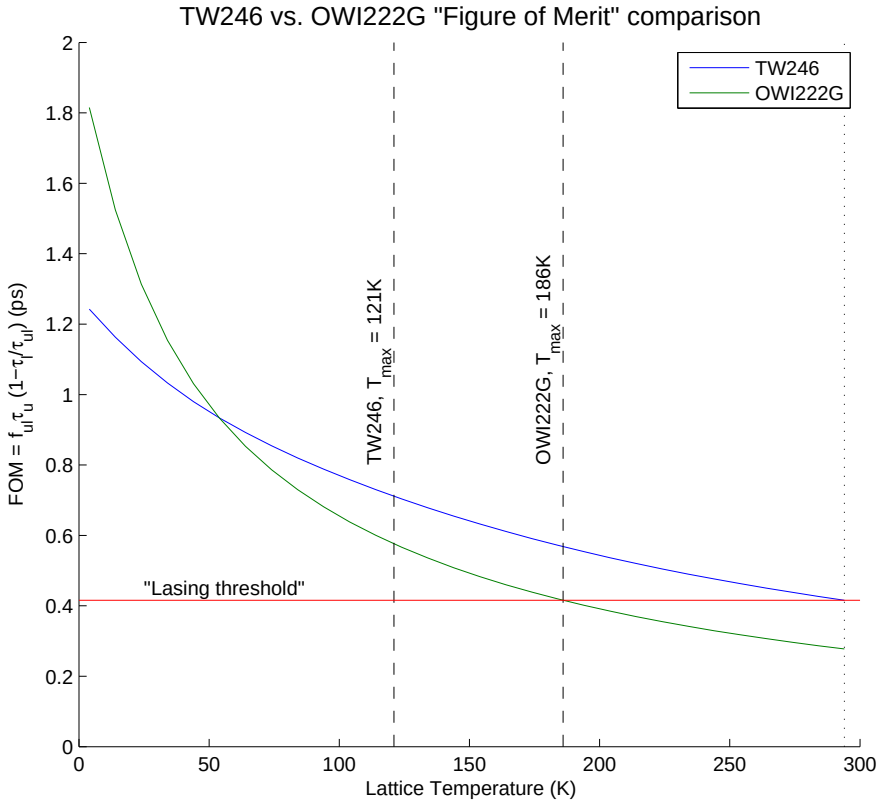


Figure 4-16: Comparison of the figure-of-merit defined in equation (2.84) between TW246 and OWI222G. Only LO phonon scattering is considered, using the thermally-activated approximation equation (2.54). The electronic temperature used in these calculations is assumed to be 100K above the lattice temperature.

The reasons for this poor performance are not known. In figure 4-15b, no value for T_0 is reported because the threshold current density does not fit well to an exponential curve at high temperatures. This is on account of an extremely steep rise in $J_{\rm th}$ as the end of lasing is approached. Moreover, $J_{\rm max}$ starts to rise prior to the end of lasing. For example, the point of peak optical power moves further behind $J_{\rm max}$ with increasing temperature.

In [63], Kumar has speculated that this $J_{\rm max}$ behavior and the performance degradation of the device might be due to hot-phonon induced absorption of upper level electrons into upper parasitic levels. This is illustrated in figure 4-17. At design bias, the upper laser level is approximately in LO phonon resonance with upper parasitic level 4. The scattering time to these parasitics therefore increases as resonance is approached, providing one possible explanation for why at higher temperatures the power peaks before $J_{\rm max}$ is reached. In figure 4-17, the scattering lifetimes from the the levels $1' - 3$ to the parasitics $2' - 4$ are in the range of $0.4{\rm ps}/N_{LO}$ to $0.8{\rm ps}/N_{LO}$, where N_{LO} is the LO phonon occupation factor. At cryogenic temperatures, $N_{LO} \sim 0$ in equilibrium, but a nonequilibrium N_{LO} could cause significant scattering.

But absent empirical evidence for the existence of hot LO phonon populations in THz QCLs, this hypothesis remains highly speculative. Empirical evidence does, however, exists for the existence of nonequilibrium interface TO phonons [64]. Computational studies employing Monte Carlo methods also support the existence of nonequilibrium LO phonon populations in THz QCLs [65, 66]; in [66], Lu and Cao report nonequilibrium occupation factors as high as 0.2 at 25K. However, the applicability of Monte Carlo methods to QCL simulation remains questionable, so such computational studies need be treated with some caution.

There is a more straightforward manner in which level 4 can act as an upper level parasitic: it could be just that level $2'$ electrons tunnel into level 4, and then relax into level 2 through LO phonon emission, and then tunnel into level 4 of the next module, and so forth. Through this path, the electrons entirely bypasses both the upper laser level and the injector.

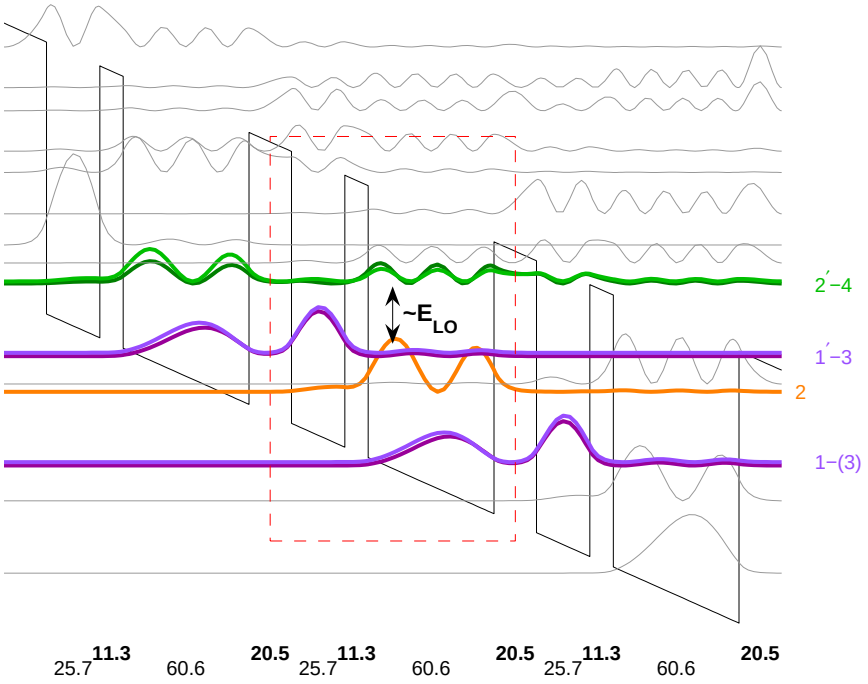


Figure 4-17: 3-module simulation of TW246 showing hypothetical mechanism of hot-phonon absorption. At the design bias, the upper laser level, depicted in purple (level 3, shown here anticrossed with the previous module's injector as $1' - 3$) is coincidentally almost in LO phonon resonance with the upper level parasitics shown in green (level 4, which is anticrossed with $2'$; calculated gap is 34.8meV). The period of one module is boxed in red.

Summary

This generation of devices produced mixed results regarding increased diagonality in design. On the one hand, OWI222G achieved a record $T_{\max}$=186K. On the other hand, TW246 performed much more poorly than OWI222G despite being an ostensibly much better design according to available theoretical tools.

The intended lasing mechanism of OWI185E-M1 also disappointed, but its unintended high-bias lasing has established scattering assisted injection as a viable design strategy that may prove superior to resonant tunneling injection.

4.2.4 Third generation designs

Overall, designs in the 3rd generation are characterized by three major changes. First, all designs were made much more diagonal than their 2nd generation predecessors, in order to increase the upper level lifetime. Second, the injection anticrossings were generally made larger, in a bid to improve current transport and hence maximum gain. Third, some designs employ a larger phonon gap or used a two phonon cascade in order to combat thermal backfilling.

This generation of devices had somewhat suspect quality of growth, consistent with our MBE collaborator, Dr. Reno, having recently moved to a new MBE machine. Several wafers exhibited unusually broad X-ray diffraction data, or showed a split period suggesting two different periodicities in the superlattice. This set of devices also largely switched from bulk doping to δ-doping.

OWI190E-M2

OWI190E-M2 is a another descendant of OWI180E. There were two growths of OWI190E-M2: VB0288 was overgrown by 3.1% and VB0287 was overgrown by 7.4%. Accounting for overgrowth in this design, the oscillator strength is reduced from 0.37 in OWI185E-M1 to 0.15 and 0.14 in the two wafers (the targeted oscillator strength was 0.21). The injector anticrossing was designed to be larger than OWI185E-M1, although overgrowth resulted in it being the same in the case of VB0288, and smaller

in the case of VB0287. The collector miniband width was also smaller in both cases. Finally, the phonon depopulation gap was increased to ∼46meV in a bid to suppress thermal backfilling. As with OWI185E-M1, the change of so many parameters at once hampers analysis. Design specifications for both wafers are presented in figure 4-18. Experimental results for VB0288 and VB0287 are presented in 4-19 and 4-20 respectively.

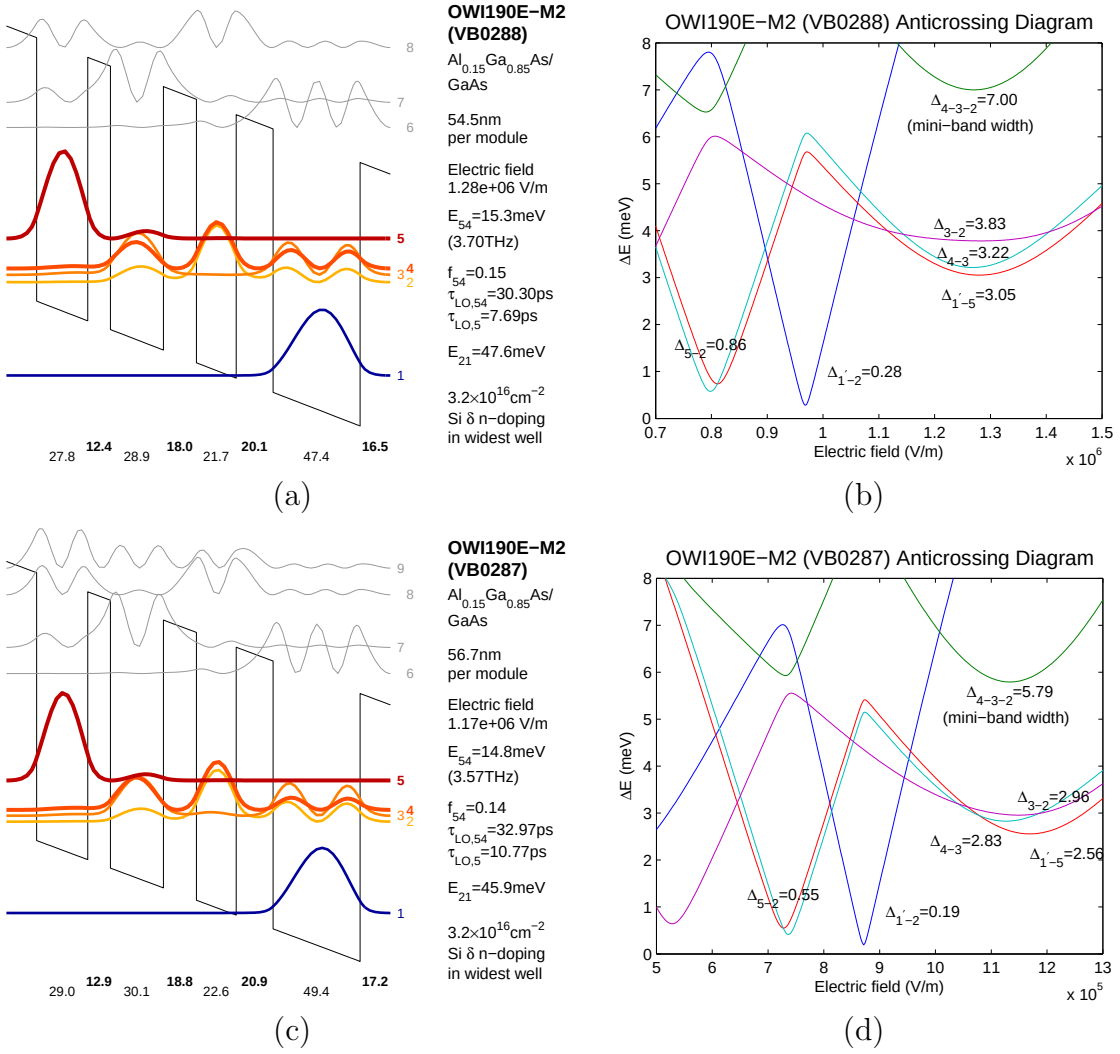


Figure 4-18: Design parameters for OWI190E-M2, VB0288 (overgrown by 3.1%) and wafers VB0287 (overgrown by 7.4%). (a) Calculated single-module band diagram at optimal injection bias for VB0288. (b) Calculated anticrossings for VB0288. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collection miniband width is Δ_{4-3-2}. Injection anticrossing is $\Delta_{1'-5}$. (c) and (d) are the same plots for VB0287.

Both VB0288 and VB0287 lased up to $T_{\max}$=137K, and showed a marked decrease

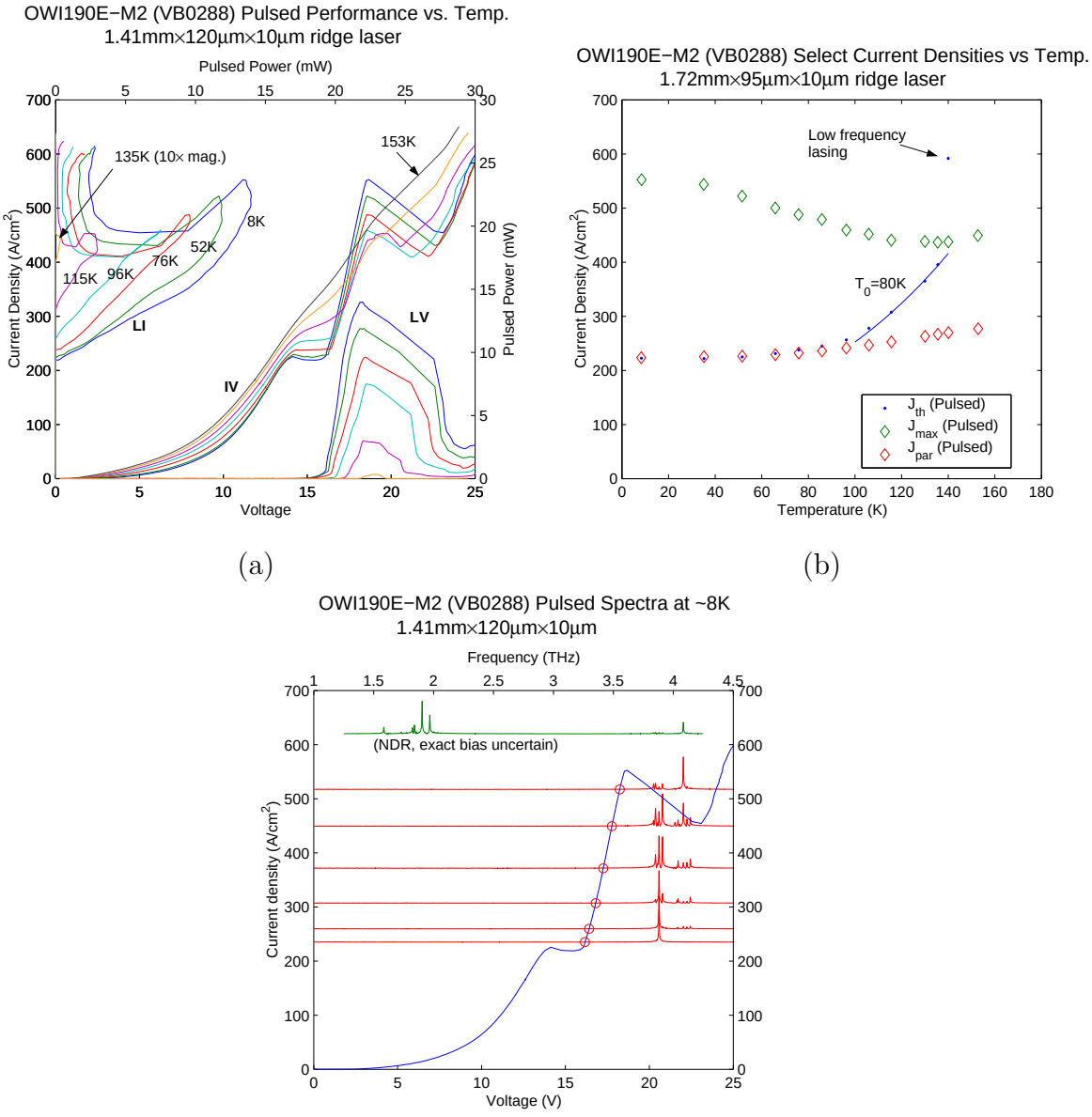


Figure 4-19: Experimental results for OWI190E-M2, wafer VB0288 (overgrown by 3.1%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 14.0V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location.

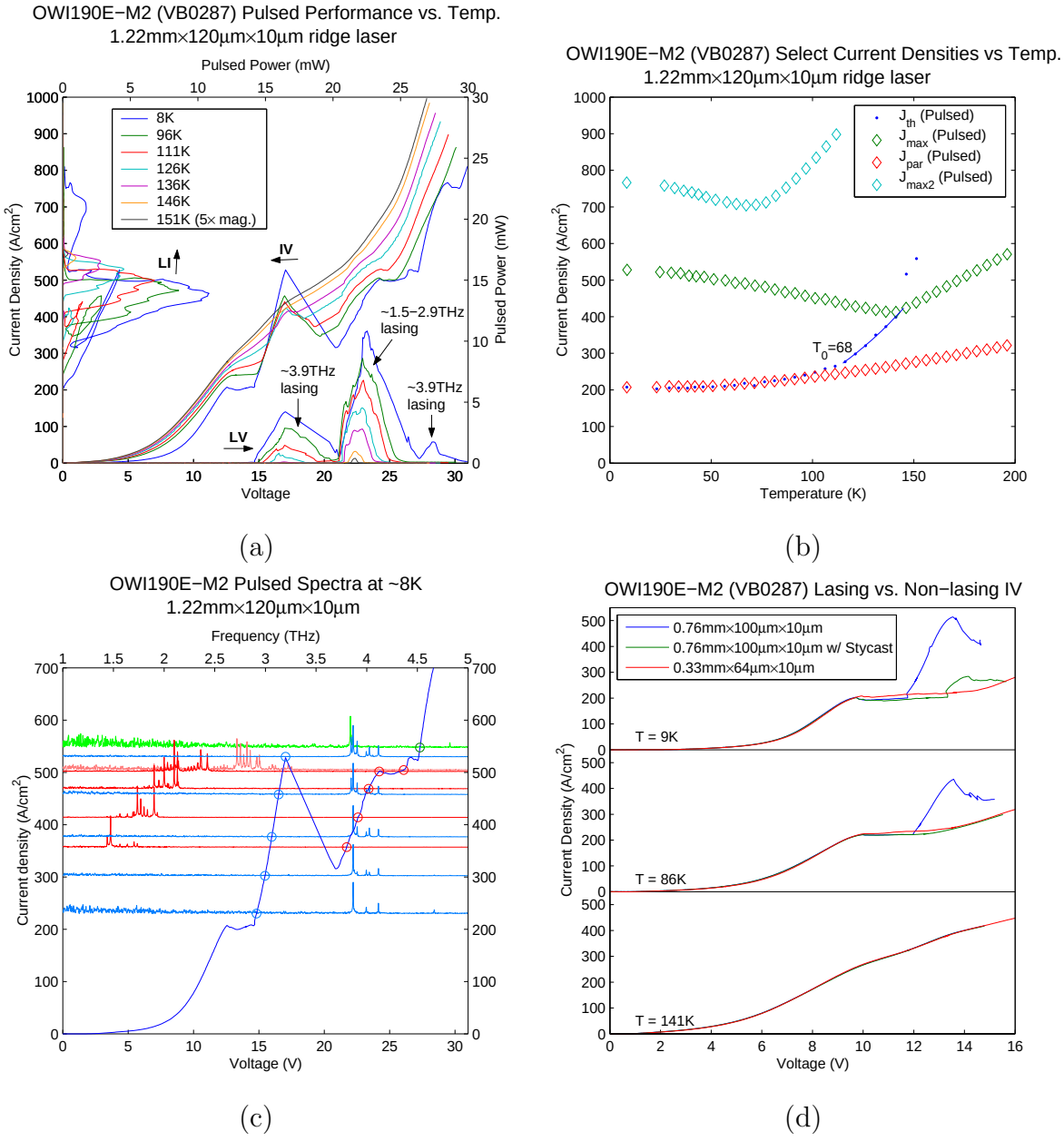


Figure 4-20: Experimental results for OWI190E-M2, wafer VB0287 (overgrown by 7.4%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold ($J_{\rm th}$), maximum ($J_{\rm max}$) and parasitic ($J_{\rm par}$) current densities versus temperature. $J_{\rm par}$ is evaluated from the IVs at 12.6V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) Lasing versus non-lasing IVs. Lasing was suppressed by covering devices with Stycast 2850 epoxy (catalyst 23LV) or by using extremely small devices. At low temperatures, the Stycasted device shows a delayed onset of lasing, and the smallest device does not lase at all. At higher temperatures, lasing ceased in all three devices, and all IV's converge. Note that all features of the IV's in d happen at lower voltages.

in $J_{\max}$ versus temperature. While this has been observed before in previous designs (eg. OWI185E-M1), the drop has never been so dramatic. The interpretation is that transport in this design is limited by the upper level lifetime, due to the large injector anticrossing and long upper level lifetime. Lasing plays a significant role in transport above threshold, and so the decrease of $J_{\max}$ is caused by diminishing lasing as temperature increases. As temperature increases, the *non-radiative* lifetime decreases, but the *radiative* lifetime increases (ie. less lasing occurs), but such that overall upper level lifetime increases with temperature while the device continues to lase. Another explanation for falling $J_{\max}$ is thermal backfilling of electrons from the injector into the collector levels, but this is inconsistent with the larger LO phonon gap of OWI190E versus OWI185E-M1—if thermal backfilling is the cause, $J_{\max}$ should have dropped more prominently with temperature in OWI185E-M1.

Additionally, OWI190E-M2 had a prominent parasitic knee in its IV characteristics. This parasitic is likely causing early NDR, and may be harming performance. Williams has identified such parasitics as a cause of poor temperature performance in previous designs (FL175C and its descendants, detailed in [46]). This suspicion was confirmed in the measurement of non-lasing devices (see figure 4-20), which demonstrated that J_{par} and $J_{\max}$ are practically identical in the absence of lasing.

Furthermore, like OWI185E-M1, OWI190E-M2 supports scattering assisted lasing, although this was only investigated thoroughly in VB0287. High temperature spectra were not recorded, so the optimal lasing band alignment is not as certain. But assuming injection is optimized around the same $3' - 7$ band alignment as OWI185E-M1, then the postulated band diagram is shown in figure 4-21 ($3' - 7$ anticrossing gap is 0.95meV). As see in figure 4-21, the 4-3 lasing transition is much more diagonal. In principle, this increased diagonality should have been beneficial, as the reduced wavefunction overlap should improve injection selectivity. This is reflected in the LO phonon scattering rates calculated in figure 4-22. In contrast to OWI185E-M1, even the 5-4 injection is substantially more selective. However, the scattering assisted transition lased only up to 152K— superior to the resonant tunneling injection in this wafer, but somewhat inferior to the scattering assisted injection lasing in OWI185E-

M1. There are several possible explanations for the poorer performance: compared to OWI185E-M1, the phonon depopulation gap is larger, and the $3' - 7$ anticrossing gap is smaller. There are probably also differences in the resonant tunneling collection by levels 3 and 2 between OWI190E-M2 and OWI185E-M1, which is optimal in neither case.

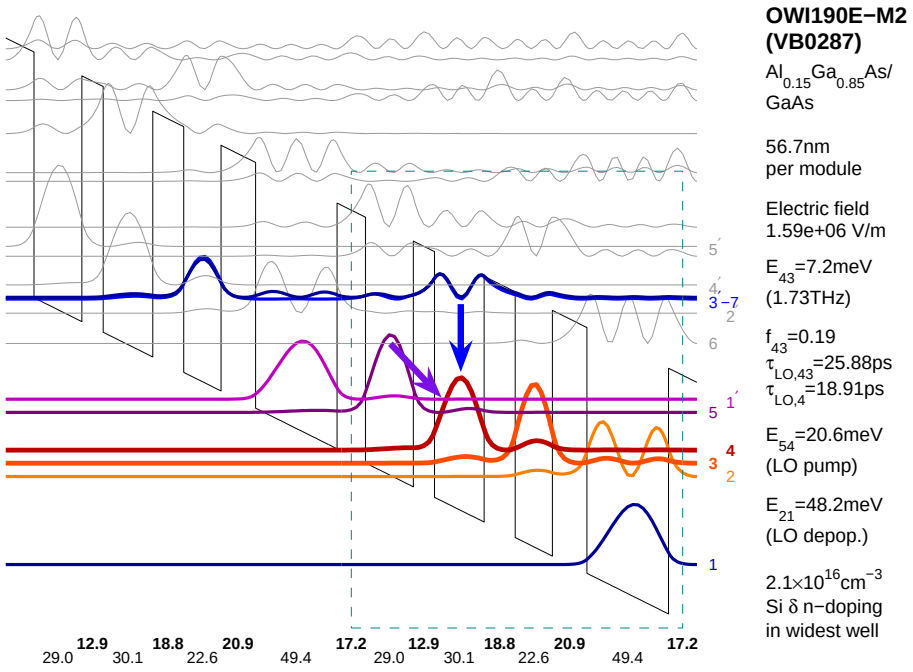
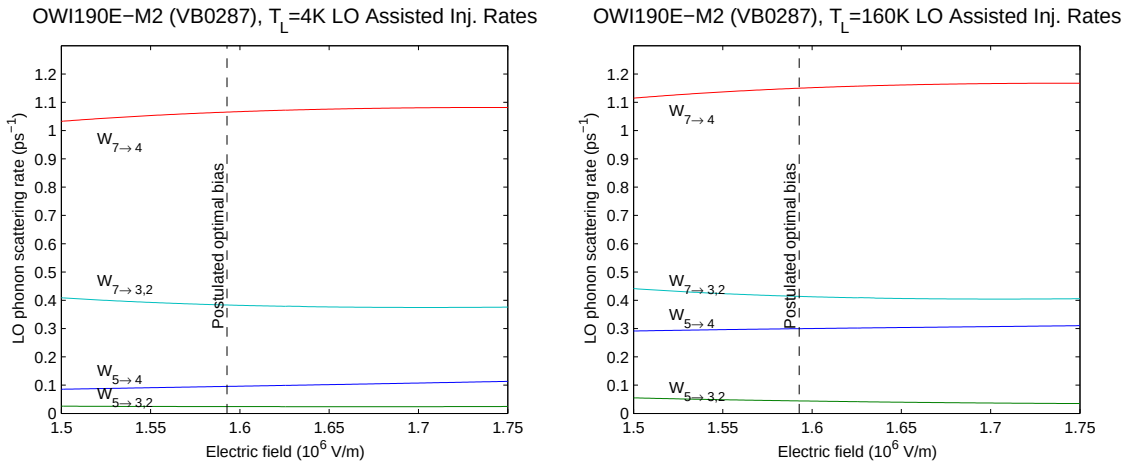


Figure 4-21: Postulated 2-module band diagram for scattering assisted lasing in OWI190E-M1, with the possible 7-4 (blue) and 5-4 (purple) scattering assisted injection. The calculated $3' - 7$ anticrossing gap is 0.95meV

OWI222G-M1

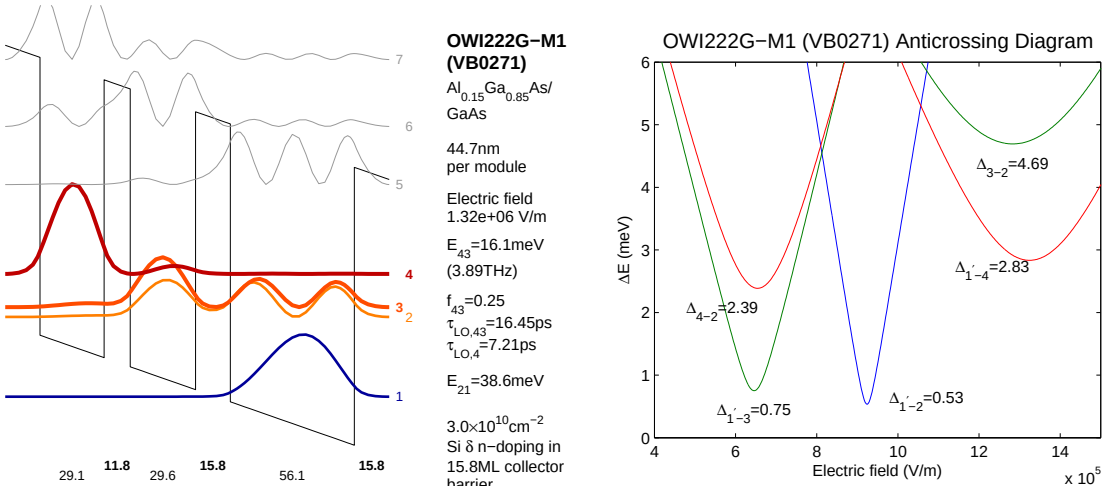
OWI222G-M1 is a modification of OWI222G. Accounting for growth inaccuracies, the oscillator strength is moderately reduced from 0.37 down to 0.25, and the injector anticrossing is increased from 2.15meV to 2.83meV. The collector anticrossing of both designs is similar (4.52meV in OWI222G, and 4.69 meV in OWI222G-M1). Design specifications and experimental data are presented in figures 4-23 and 4-24 respectively.

Like OWI190E-M2, OWI222G-M1 exhibited a falling $J_{\max}$ with rising tempera-



(a) (b)

Figure 4-22: LO phonon scattering rates in OWI190E-M2 vs. electrical bias for 7-4 injection and 5-4 injection at (a) 4K (b) 160K. Rates are calculated based on a 1-module band diagram simulation. Electronic temperature is assumed to be 100K above the lattice temperature.



(a) (b)

Figure 4-23: Design parameters for OWI222G-M1, wafer VB0271 (undergrown by 1.5%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}, and injection anticrossing is $\Delta_{1'-4}$.

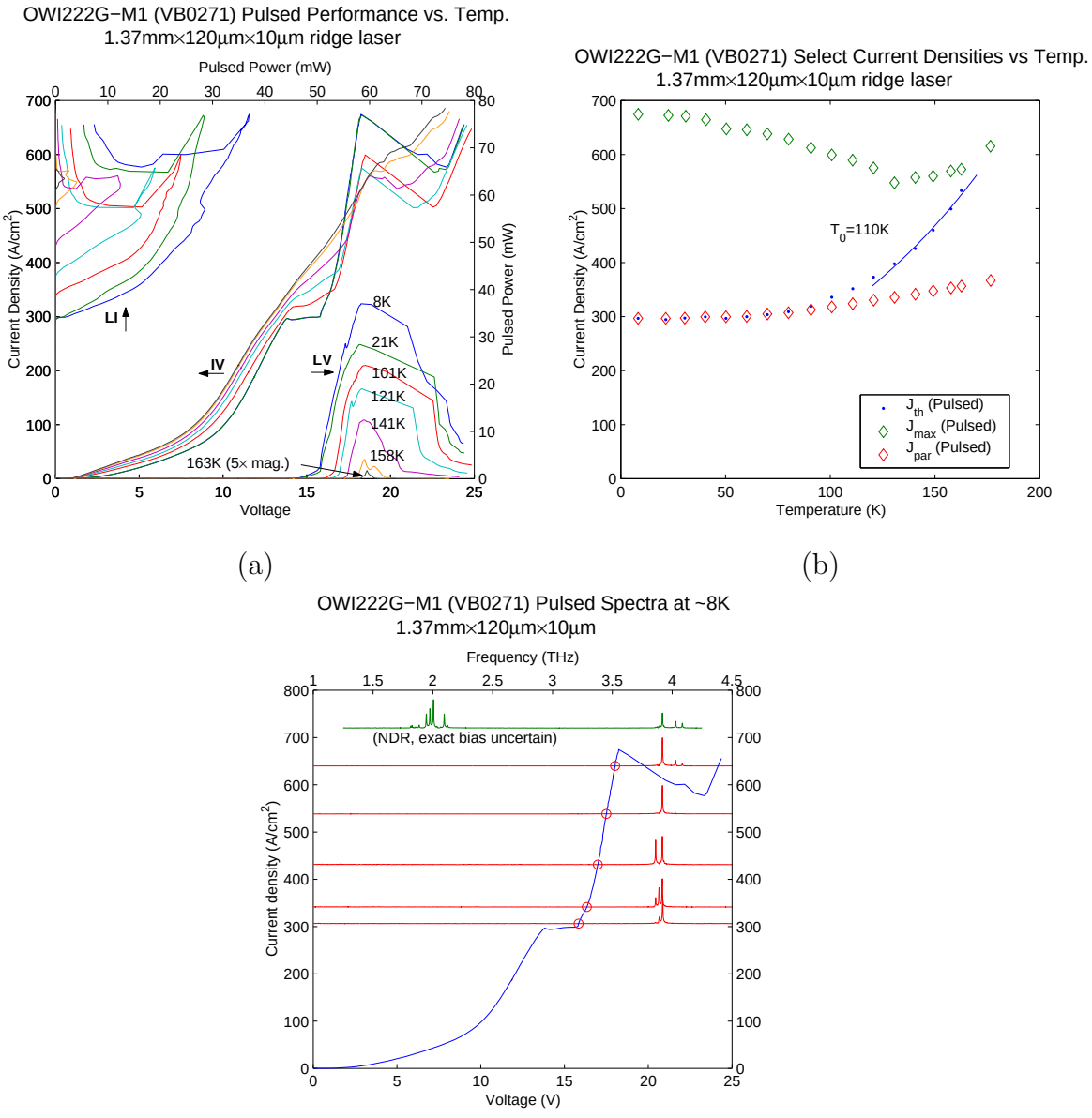


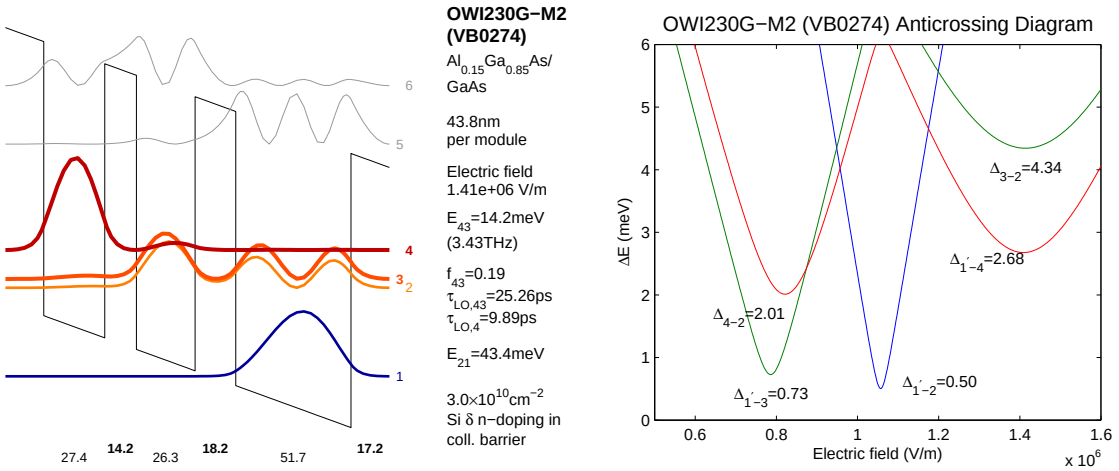
Figure 4-24: Experimental results for OWI222G-M1, wafer VB0271 (undergrown by 1.5%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 13.5V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) Lasing versus Non-lasing IV measurement at ∼8K.

ture, although unlike in OWI190E-M2, this trend does not persist all the way until the end of lasing (see figure 4-24b). It also seemingly suffers more from subthreshold parasitics than OWI222G, despite having a nearly identical parasitic anticrossing (0.53meV for OWI222G-M1 versus 0.50 meV for OWI222G); whereas OWI222G reaches threshold without NDR, OWI222G-M1 has an early NDR.

OWI230G-M2

OWI230G-M2 is a modification of OWI222G. Compared to OWI222G-M1, it is even more diagonal (oscillator strength 0.19), and also possesses an even larger injector anticrossing (2.68meV). In addition, to test whether thermal backfilling is an issue or not, the phonon depopulation gap, E_{21}, was increased to 43.4meV. Design specifications and experimental data are presented in figures 4-25 and 4-26 respectively.

The wafer for this device, VB0274, is strongly suspected to have defective growth on account of its abnormal IV characteristics, and lased only up to 88K. As such, it admits few conclusions.



(a) (b)

Figure 4-25: Design parameters for OWI230G-M2, wafer VB0271. Ostensibly undergrown by 1.5%, but suspected of having more serious growth problems. (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}, and injection anticrossing is $\Delta_{1'-4}$.

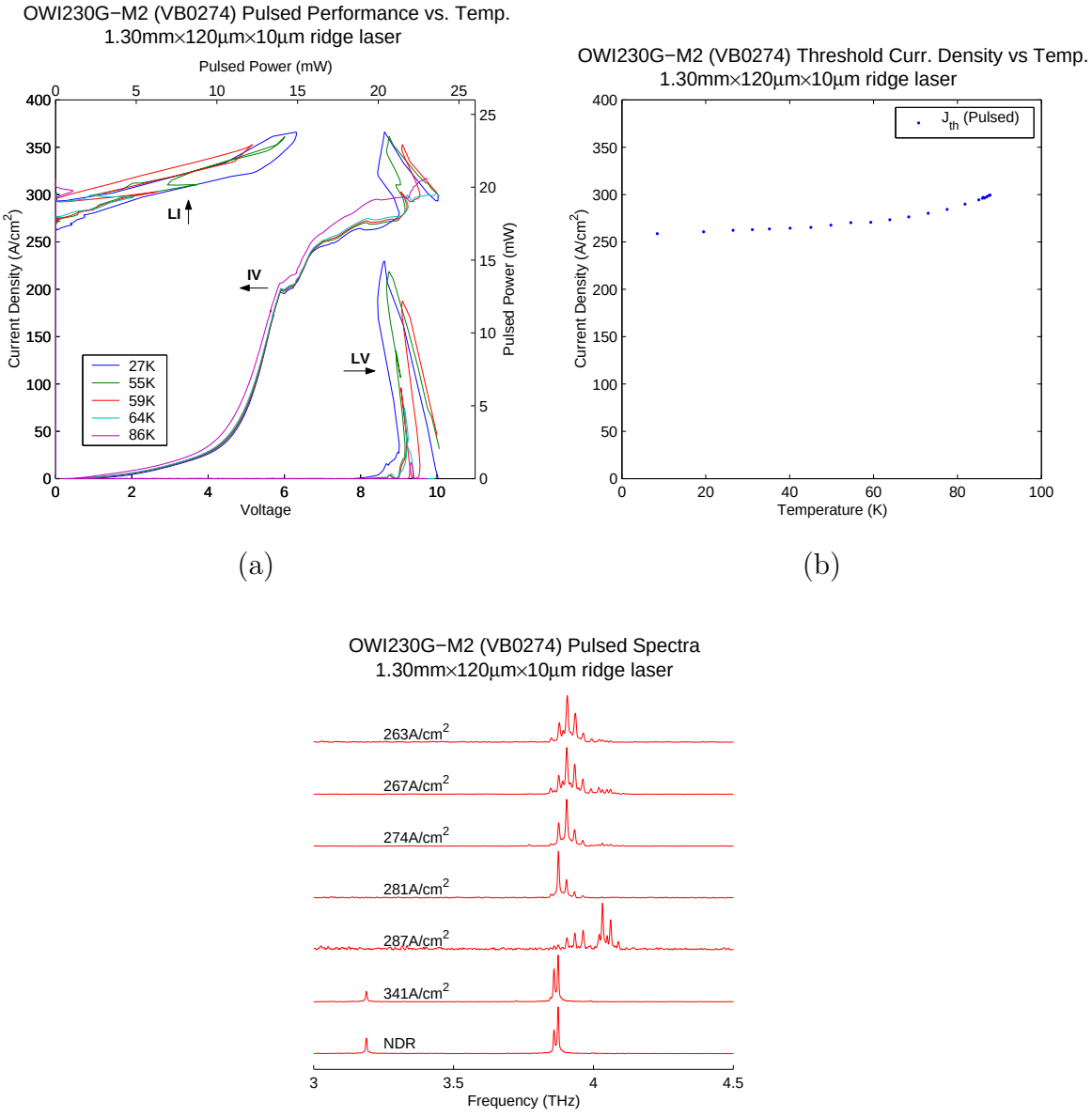


Figure 4-26: Experimental results for OWI230G-M2, wafer VB0271. Ostensibly undergrown by 1.5%, but suspected of having more serious growth problems. (a) Pulsed LIV versus temperature measured for a wide device. The highly irregular IV suggests questionable crystal growth. (b) Threshold (J_{th}) current density versus temperature. (c) Pulsed spectra at ∼8K.

OWI235G-M3

OWI235G-M3 is a modification of OWI222G. It is the most diagonal of the OWI222G descendants in this generation, with an oscillator strength of 0.16. Like OWI222G-M1 and OWI230G-M2, its 2.93meV injector anticrossing is larger than OWI222G. Notably, it also had an unusually large phonon gap, $E_{21} = 55.6$meV. Design specifications and experimental data are presented in figures 4-27 and 4-28 respectively.

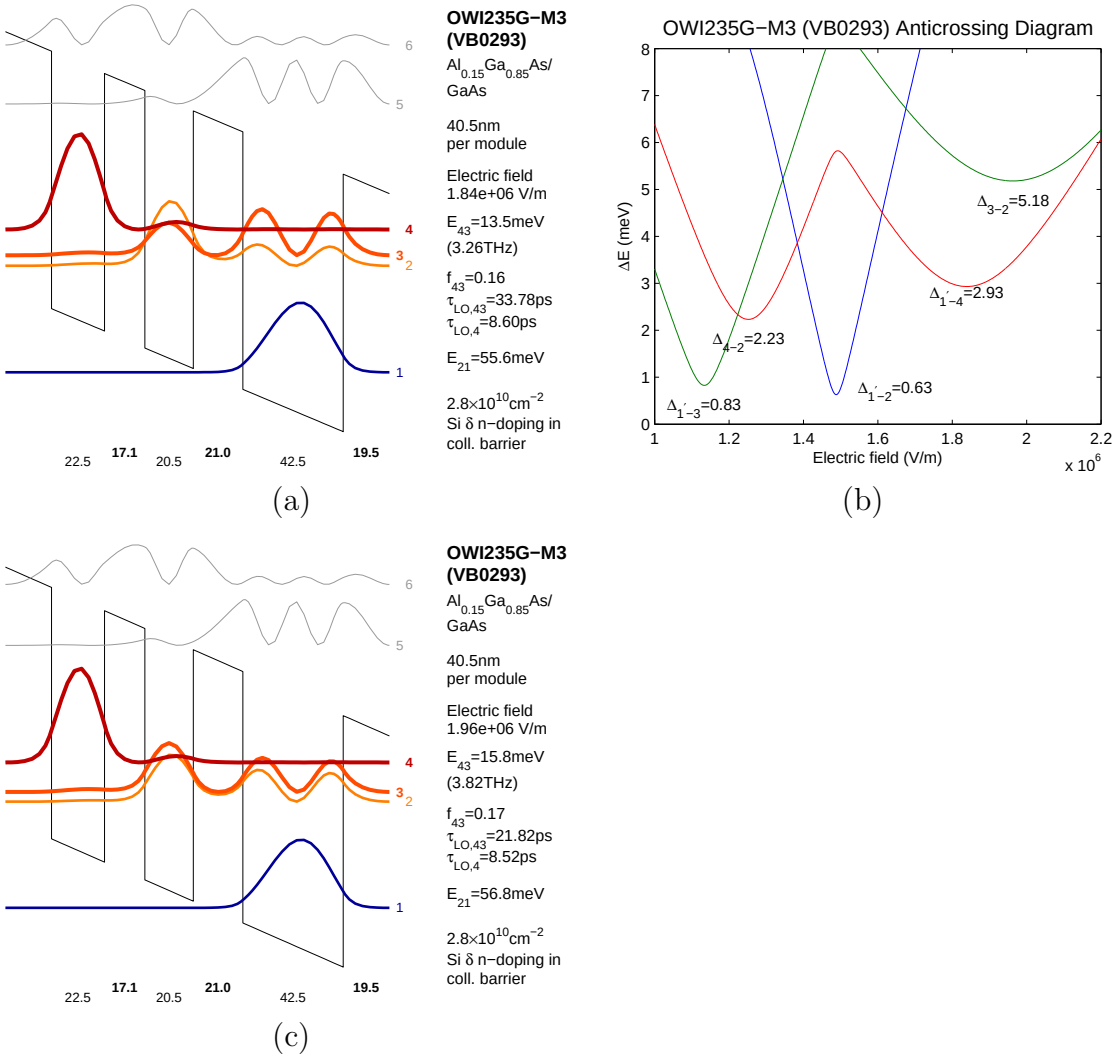
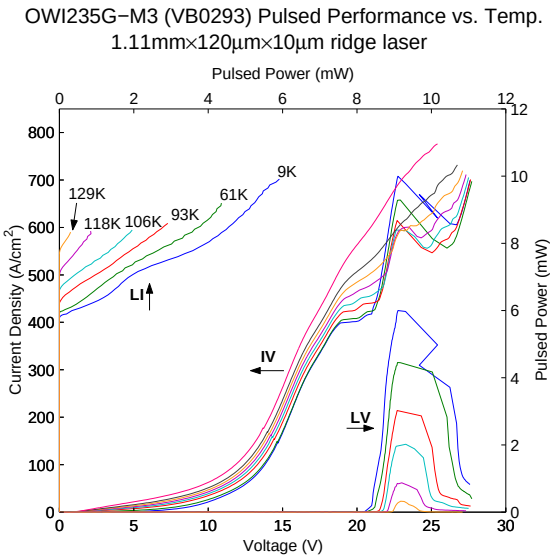
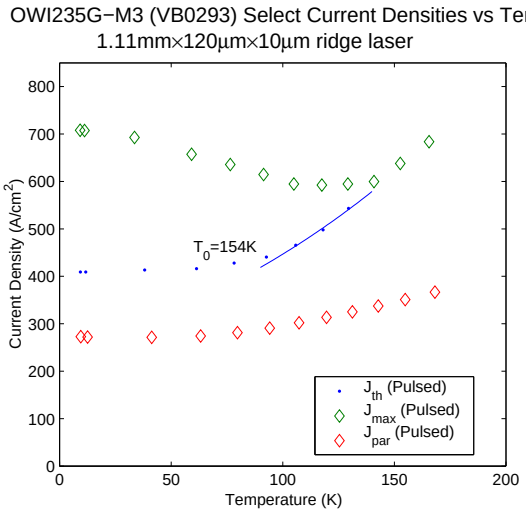


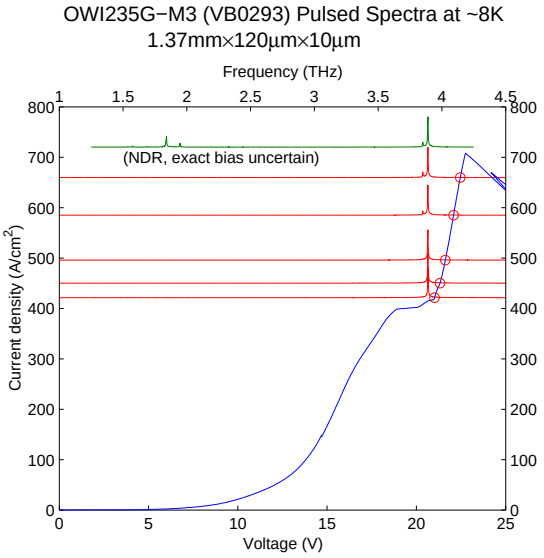
Figure 4-27: Design parameters for OWI235G-M3, wafer VB0293 (undergrown by 2.26%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}, and injection anticrossing is $\Delta_{1'-4}$. (c) Calculated single-module band diagram at optimal collection bias. This bias better reproduces experimentally observed lasing frequency.



(a)



(b)



(c)

Figure 4-28: Experimental results for OWI235G-M3, wafer VB0293 (undergrown by 2.26%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 16.4V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location.

Like OWI190E-M2 and OWI222G-M1, OWI235G-M3 exhibited a distinct drop in $J_{\max}$ with temperature. However, the IV for OWI235G-M3 is also unusual in another way. At first glance, there appears to be a prominent parasitic knee in the manner of OWI190E-M2, but closer inspection reveals there to be more structure. In fact the knee in the IV looks remarkably like the non-lasing IV of OWI222G in figure 4-12. Figure 4-29 shows non-lasing IV data for OWI235G-M3, alongside non-lasing IV data for OWI222G reproduced from figure 4-12 for comparison, labeled with the postulated location of the $1' - 2$ parasitic.

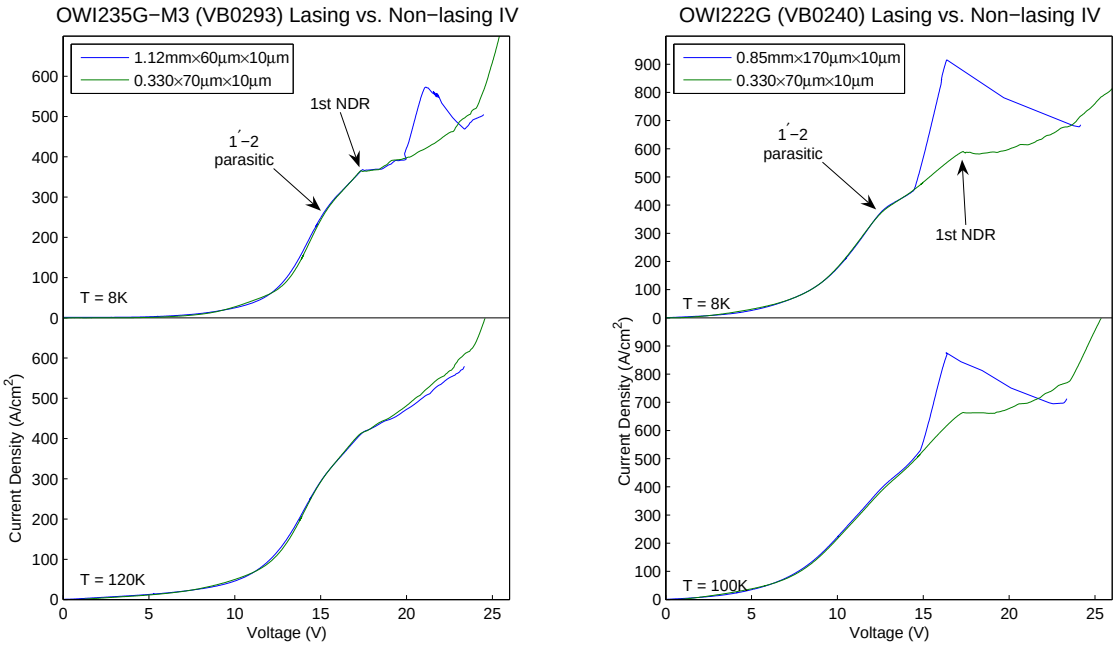
This may be explained as follows. Figure 4-27b shows that the collection anticrossing occurs after the injection anticrossing, whereas the reverse is true in most designs. The observed lasing frequency is also much higher than the calculated frequency at optimal injection bias in figure 4-27a, and a better match is obtained with the band diagram calculated at optimal collection bias in figure 4-27c. This explains why lasing occurs after the first NDR, which is caused by injector detuning.

Although this hypothesis is reasonable, it raises another question: why does the gain peak closer to optimal collection rather than optimal injection? This suggests that the lower level lifetime may be limited by the resonant tunneling collection process. Although this situation may be unique to OWI235G-M3, one cannot help but remark that the DSL family of devices, designed to combat this specific problem, have empirically performed worse than standard OWI and FL designs.

TW260-M1

TW260-M1 is a modification of TW246. A large number of parameter changes from its predecessor hinders the interpretation of the experimental data. The oscillator strength was reduced from 0.52 to 0.29. The injection anticrossing was increased from 1.67meV to 2.26meV. The phonon gap was increased from 37.2meV to 51.7meV. The lasing frequency was reduced from 4.6THz to 3.8THz (granted, this frequency reduction was necessary to accommodate the larger phonon gap).

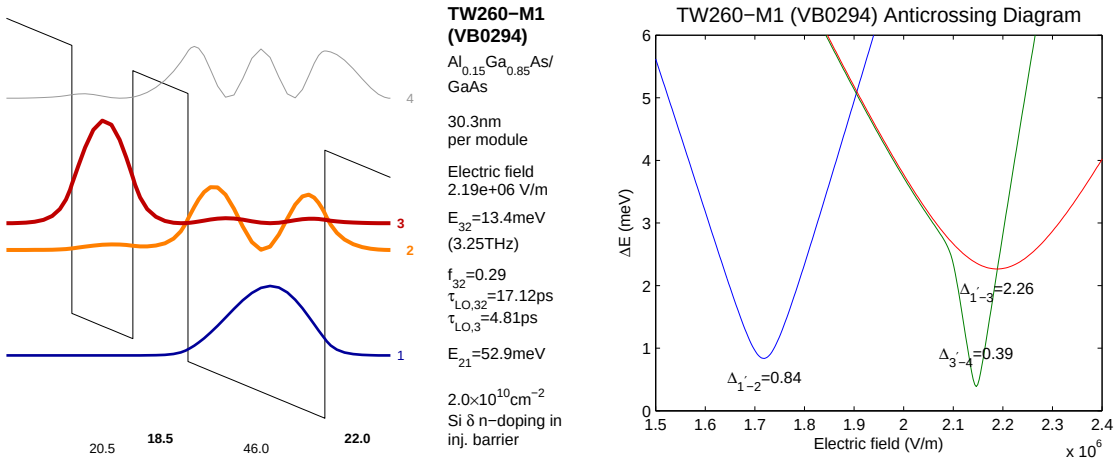
This device did very poorly, lasing only up to 43K. The reasons for this poor performance are uncertain. From preceding data on large phonon gap designs (OWI190E-



(a) (b)

Figure 4-29: Comparison of OWI235G-M3 and OWI222G non-lasing IVs. Lasing in OWI235G-M3 starts after the 1st NDR of its non-lasing IV, whereas in OWI222G it starts between the $1' - 2$ parasitic and the 1st NDR.

M2, OWI235G-M3), thermal backfilling is unlikely to be limiting performance, so the larger phonon gap may have hindered lower level depopulation. Another strong possibility is that leakage into continuum may have been a problem, as the anticrossing diagram of figure 4-30b shows the upper laser level $3'$ to be anticrossed with the pseudo-continuum level 4 in the vicinity of optimal injection bias. And then there remains the lingering possibility that it might be diagonality itself that is damaging performance, however depressing that notion might be.



(a) (b)

Figure 4-30: Design parameters for TW260-M1, wafer VB0294 (overgrown by 0.1%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Injection anticrossing is $\Delta_{1'-3}$. As this is a DSL design, there is no collection anticrossing.

TP3W200

"TP" stands for "two phonon." The TP devices were designed to combat thermal backfilling, using two back-to-back intrawell phonon emissions to depopulate the lower laser level.

TP3W200 is an OWI family design. It is essentially a TP variant of OWI222G, employing a one injector well and a two well active region. Accounting for overgrowth, it was designed to lase at 2.7THz with an oscillator strength of 0.19. The 2nd excited state of the widest quantum well is used as the collector, and the two-phonon cascade

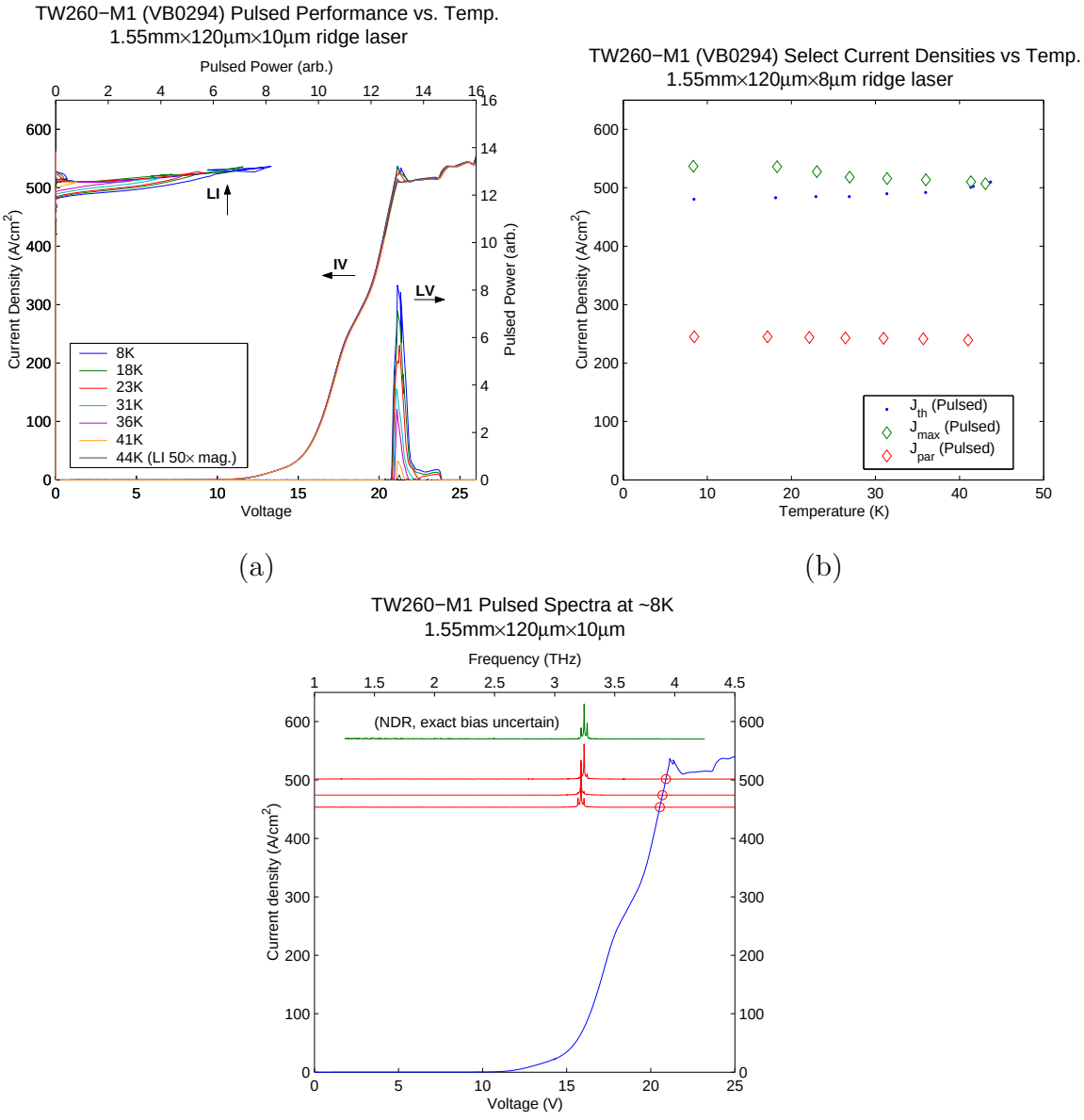


Figure 4-31: Experimental results for TW260-M1, wafer VB0294 (overgrown by 0.1%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold ($J_{\rm th}$), maximum ($J_{\rm max}$) and parasitic ($J_{\rm par}$) current densities versus temperature. $J_{\rm par}$ is evaluated from the IVs at 17.6V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location.

has energy level spacings E_{32}=35.8meV and E_{31}=62.4meV. Design specifications and experimental data are shown in figures 4-32 and 4-33 respectively.

TP3W200 exhibited a strong parasitic knee in its IV, but lacked the characteristic plateau of the single phonon designs. Interestingly, TP3W200 ended up being a two color device. The designed 5-4 transition lased only weakly, and lased only to up $T_{\max}$=54K. However, at higher biases, the devices lased at the much higher frequency of 3.9THz, also exhibiting a strong Stark shift to higher frequencies with increasing biases (as high as 4.3THz). This high frequency lasing was substantially higher in power and more robust, lasing up to $T_{\max}$=96K. This suggest a scattering assisted injection mechanism in the manner of OWI185E-M1. (Note that in figure 4-33, only (d) shows the data for the $T_{\max}$ = 54K/96K device. (a)-(c) show more complete data taken on a different device from the same fabrication).

The spectral data in figures 4-33c and 4-33d are ambiguous, showing resonant-phonon lasing at ∼3.2THz and ∼2.95THz respectively, but in both cases the lasing occurs at higher than expected frequencies. Similar to OWI235G-M3, a better match to the experimentally observed lasing frequencies is obtained by calculating the band diagram at optimal collection bias rather than optimal injection bias, as shown in figure 4-32c, suggesting that lower laser level depopulation may have been a problem. Unlike OWI235G-M3, however, there is no clear sign of NDR before threshold.

TP4W160

As TP3W200 is a TP variant of OWI222G, TP4W160 is a TP variant of OWI180E. TP4W160 had an unusual X-ray diffraction pattern showing two distinct sets of peaks of different heights. The periodicity of the minor peaks indicated 0.03% overgrowth. The periodicity of the major peaks indicated 2.2% undergrowth. The following discussion assumes 2.2% undergrowth, although the likely problems of growth must be kept in mind.

Accounting for undergrowth, TP4W160 was designed to lase at 3.14THz with an oscillator strength of 0.18. The two-phonon cascade has energy level spacings E_{32}=31.0meV and E_{31}=55.7meV. The injector anticrossing is 2.25meV (compare

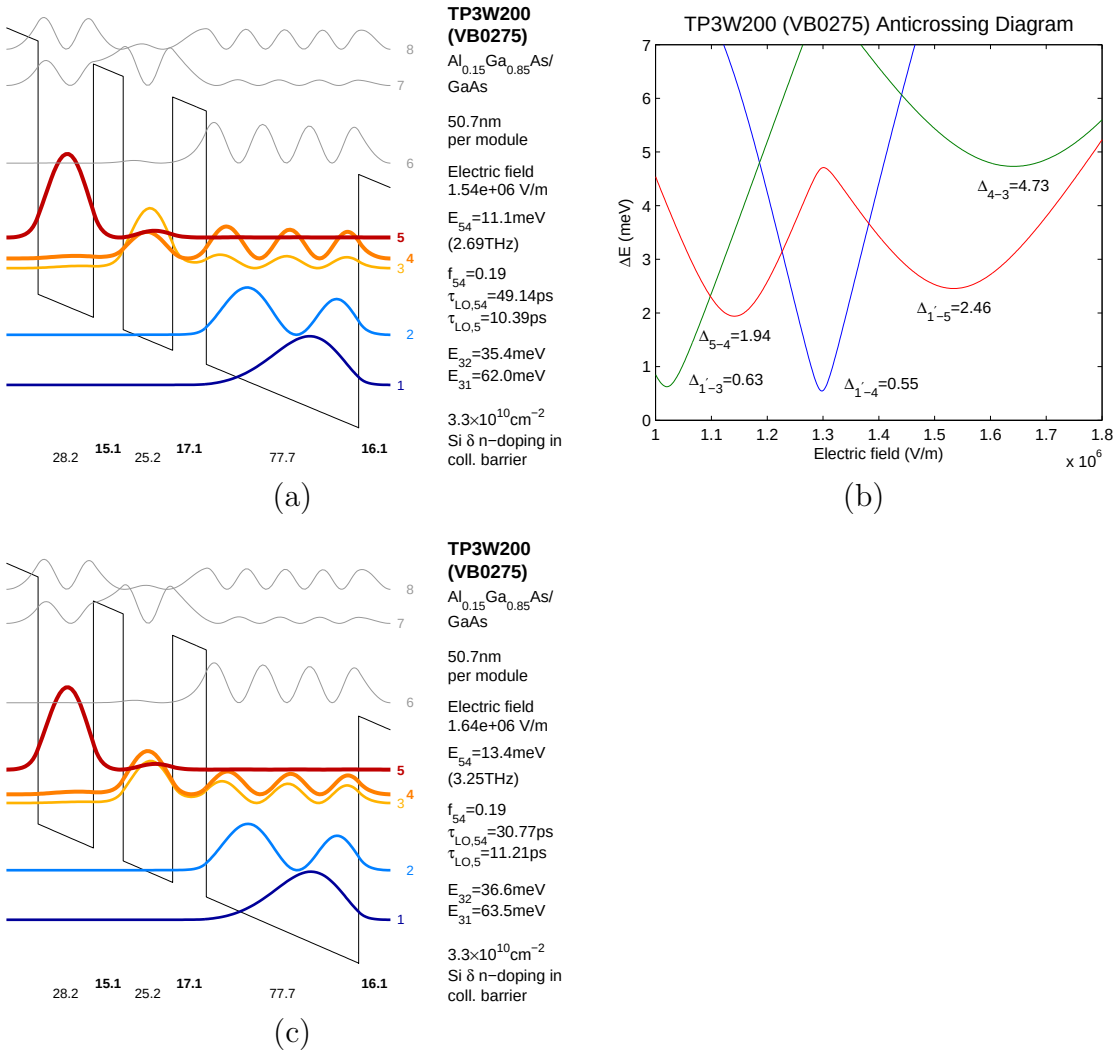


Figure 4-32: Design parameters for TP3W200, wafer VB0275 (overgrown by 0.87%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-4}$ (normally this would be denoted $1'-3$ by our naming convention, but the collector anticrossing happens after injection anticrossing in this case). The other lower level parasitic, $1'-2$ has negligible anticrossing gap. Injection anticrossing is $\Delta_{1'-5}$, and collector anticrossing is Δ_{4-3}. (c) Calculated single-module band diagram at optimal collection bias.

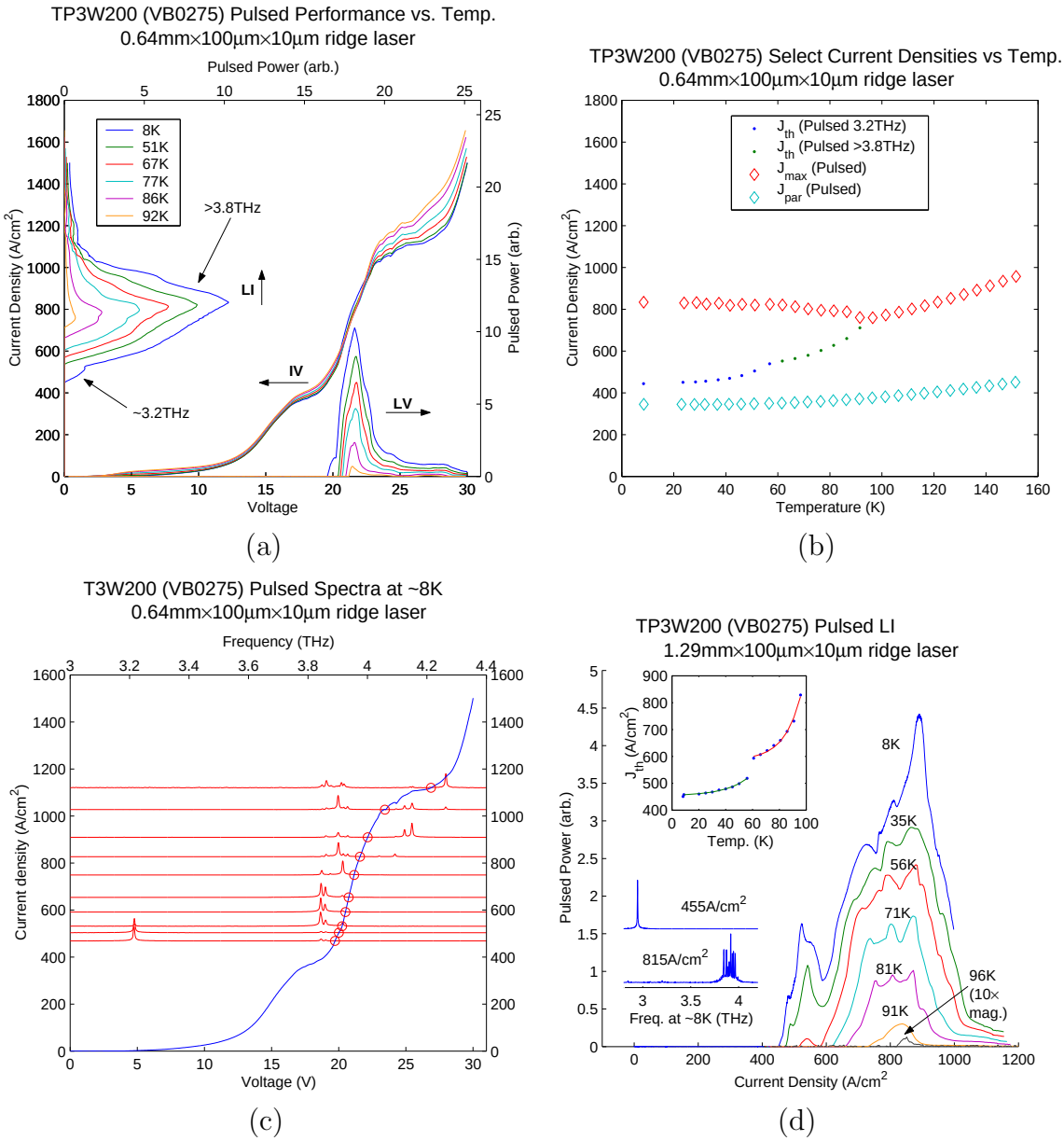


Figure 4-33: Experimental results for TP3W200, wafer VB0275 (overgrown by 0.1%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold ($J_{\rm th}$), maximum ($J_{\rm max}$) and parasitic ($J_{\rm par}$) current densities versus temperature. $J_{\rm par}$ is evaluated from the IVs at 17.0V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) Pulsed LI and spectra for another TP3W200 device that better demonstrates the two-color characteristic; no corresponding IV was taken.

2.01meV for OWI180E).

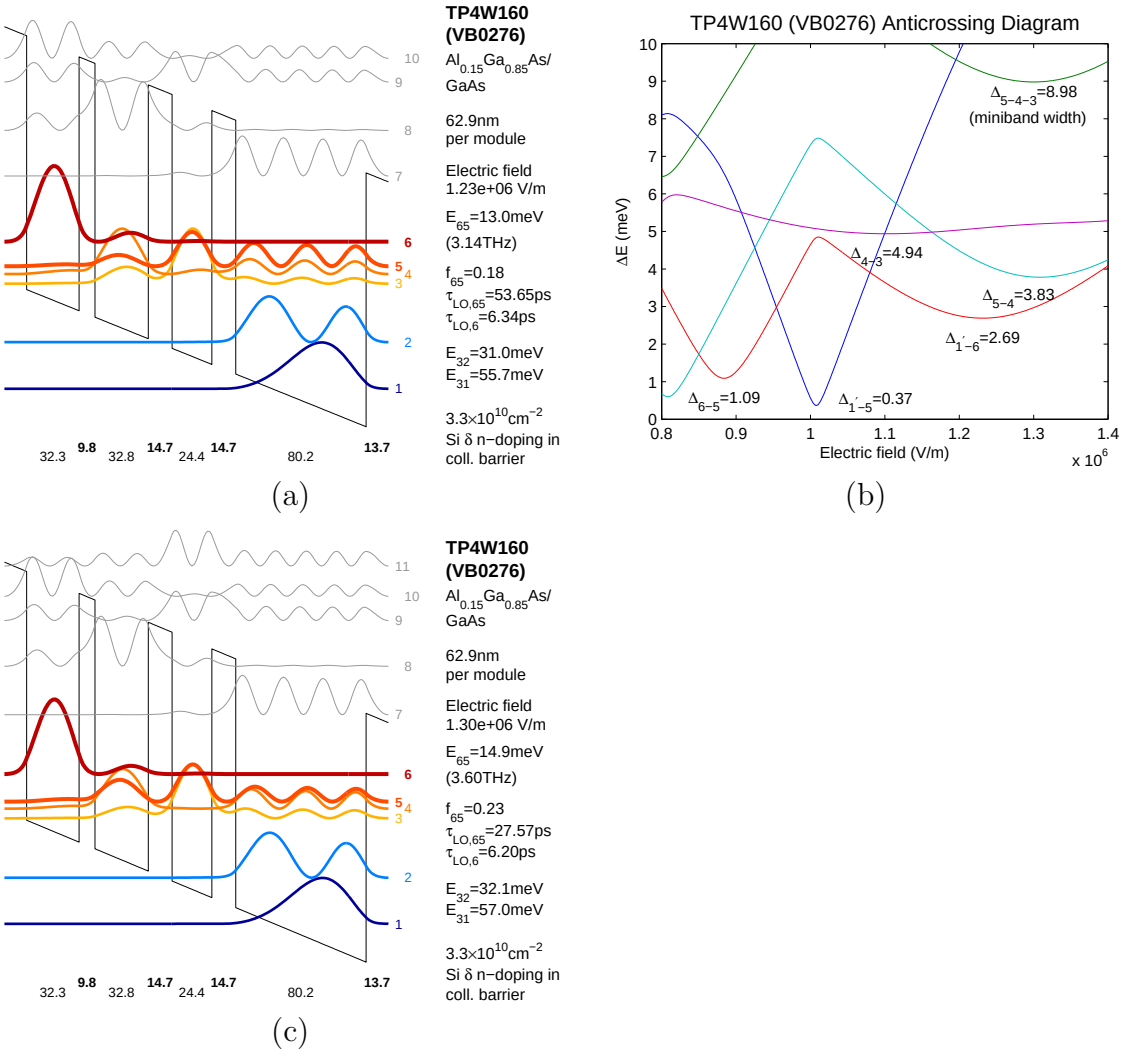


Figure 4-34: Design parameters for TP4W160, wafer VB0276 (taken as undergrown by 2.2%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-5}$. Collector miniband minimum is Δ_{5-4-3}. (c) Calculated band diagram at optimal collection bias. This agrees better with the experimental lasing frequency.

In contrast to the majority of designs in this generation, the IV exhibited only a weak knee in its IV. Experimentally, the device (wafer VB0276) lased centered around 3.9THz, much higher than expected. The lasing spectrum was also quite broad, with signs of lasing from ∼3.8THz all the way up to ∼4.2THz. TP4W160 achieved $T_{\max}$=105K. Similar to OWI235G-M3 and TP3W200, better agreement with experiment is obtained if the band diagram is calculated at optimum collection bias

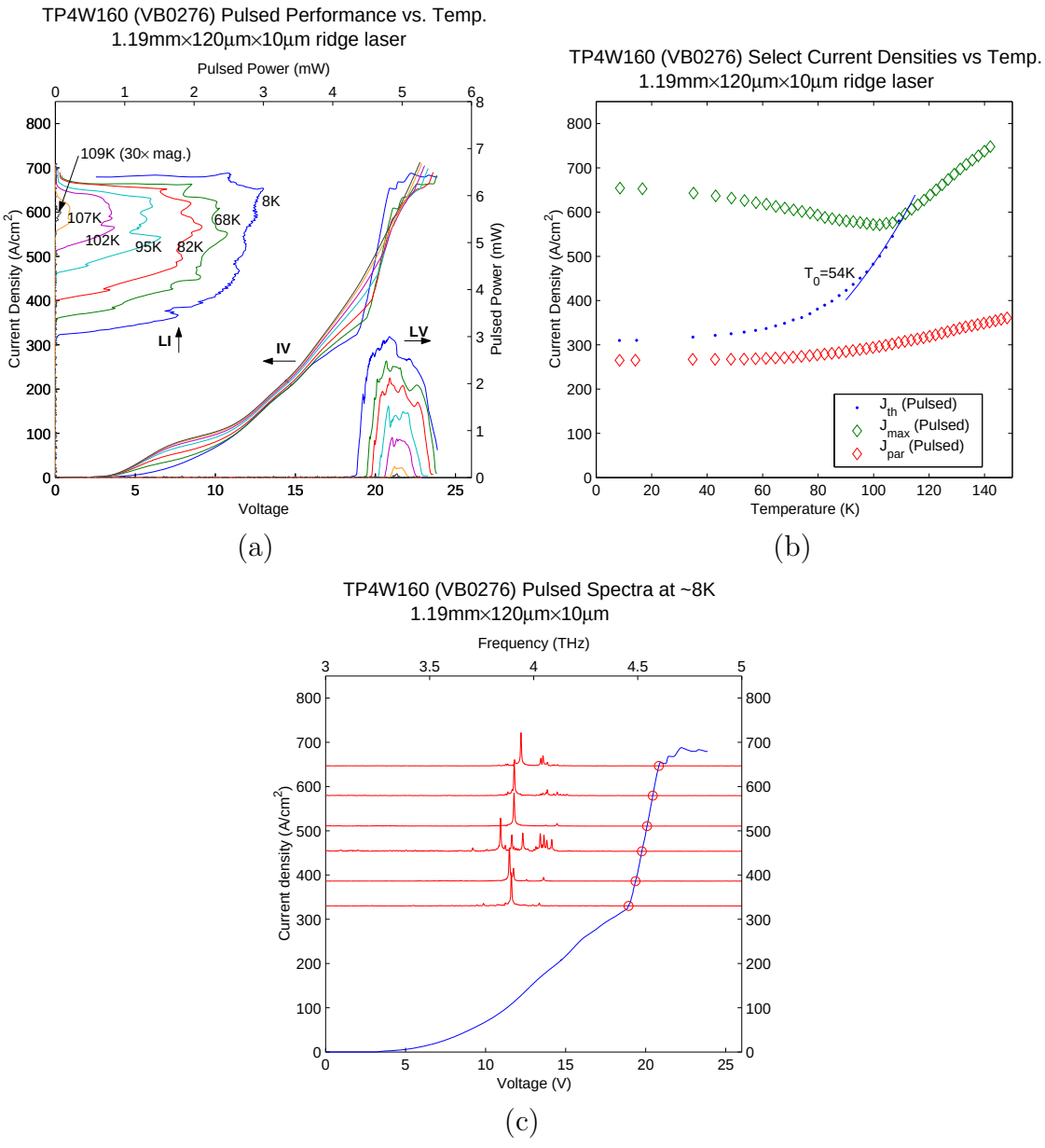


Figure 4-35: Experimental results for TP4W160, wafer VB0276 (taken as undergrown by 2.2%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 16.4V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location.

rather than optimal injection bias, suggesting slow depopulation of the lower laser level may have been an issue (although one notes that the agreement with experimental lasing frequency is still poor.)

FL190S-M2

FL190S-M2, a FL family design with a 2-well active region and 2-well injector, is a modification of FL183S. The oscillator strength is reduced from 0.77 to 0.25. In addition, the injector anticrossing is increased from 2.47meV to 2.93meV, and the collector anticrossing reduced from 5.05meV to 4.5meV.

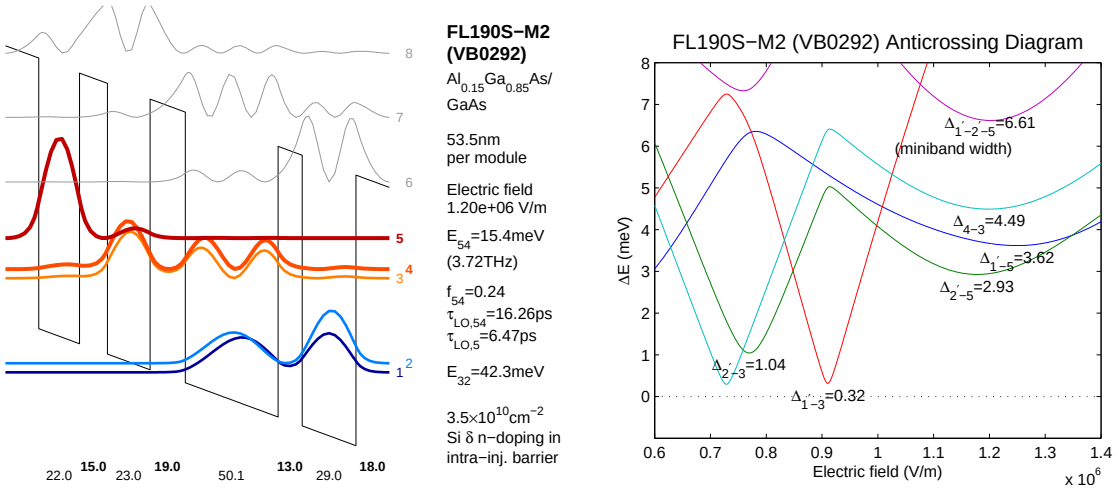
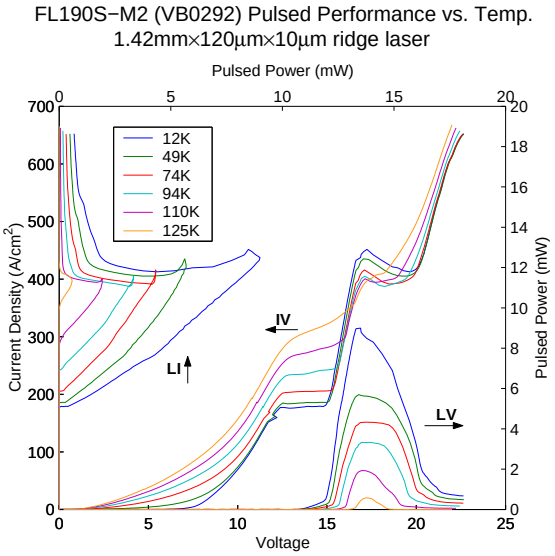


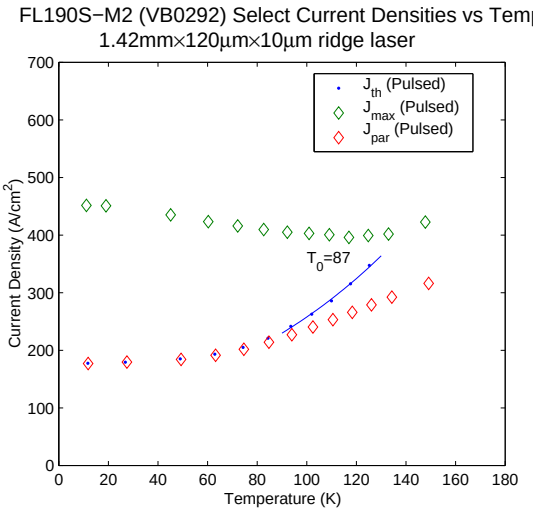
Figure 4-36: Design parameters for FL190S-M2, wafer VB0292 (overgrown by 0.16%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitics in IV are attributed to $\Delta_{1'-3}$ and possibly $\Delta_{2'-3}$. Injector anticrossings are $\Delta_{2'-5}$ and $\Delta_{1'-5}$, with overall injection miniband width given by $\Delta_{1'-2'-5}$. Collector anticrossing is Δ_{4-3}.

Like OWI190E-M2, there is a prominent knee in the IV with a flat plateau just before threshold, clearly indicating early NDR. Like other devices of this generation, it also had a characteristic drop in $J_{\max}$ versus temperature.

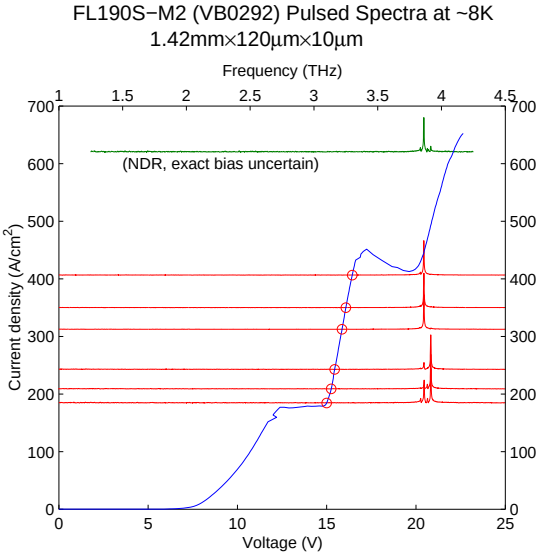
FL190S-M2 lased up to $T_{\max}$=137K. For comparison, the highest $T_{\max}$ achieved by its predecessor, FL183S, is 174K.



(a)



(b)



(c)

Figure 4-37: Experimental results for FL190S-M2, wafer VB0292 (overgrown by 0.16%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold ($J_{\rm th}$), maximum ($J_{\rm max}$) and parasitic ($J_{\rm par}$) current densities versus temperature. $J_{\rm par}$ is evaluated from the IVs at 12.5V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location.

Summary

Perhaps most importantly, 3rd generation devices demonstrated that lasers with such low oscillator strengths can still lase. This demonstrates that high diagonality is still a feasible design strategy, despite the lack-luster temperature performance of this generation.

Nearly all devices suffer from a strong parasitic knee in their IV characteristics. This knee appears to be causing early NDR, such that most devices are lasing in deep NDR, where the bias is likely inhomogeneous and hence only some fraction of the modules are contributing to lasing. If this is the case, then this generation of devices is promising in that they performed reasonably well despite the disadvantage of operating in NDR.

This generation of devices is poorly controlled for testing the importance of thermal backfilling, but the following observations can be made:

- Many devices in this generation exhibited a sharp drop in $J_{\max}$ with temperature when lasing. This can be attributed either to thermal backfilling or upper level lifetime limited transport owing to some combination of greater diagonality and larger injection anticrossings.

- This motivates the following comparisons: OWI190E-M2, OWI222G-M1, and OWI235G-M3 all exhibited pronounced drops in above-threshold current, but while they were all more diagonal than their predecessors, only OWI190E-M2 and OWI235G-M3 employed larger phonon gaps. This suggests that the falling $J_{\max}$ is symptomatic of high diagonality rather than the size of the phonon depopulation gap.

- This is consistent with comparisons between OWI190E-M2 and OWI235G-M3 and their predecessors, OWI180E and OWI222G respectively: if thermal backfilling was the cause rather than diagonality, then OWI180E and OWI222G should have exhibited more dramatic current drops on account of having a smaller E_{21}.

- OWI222G-M1 and OWI235G-M3 are similar in parameters except for their phonon gap and oscillator strength, but OWI222G-M1 clearly had a superior $T_{\max}$. As discussed up to this point, diagonality should not *hurt* performance (admittedly, however, theory and experiment in this field have a long history of divergence). If one assumes that the decreased oscillator strength of OWI235G-M3 did not hurt its temperature performance, then the decreased temperature performance is likely due to its larger E_{21}.

- Neither of the two working two-phonon designs performed well at all.

The above considerations cannot definitively rule out thermal backfilling as a barrier to high temperature performance, but it does appear that all available evidence points to the contrary.

Negligible thermal backfilling implies that the decrease in $J_{\max}$ versus temperature likely reflects transport dominated by the upper laser level lifetime, ie. $|\Omega| \gg 1/\sqrt{\tau\tau_{||}}$. However, the results of this generation do not clearly tell whether devices are operating in this regime solely because of increased anticrossings (larger $|\Omega|$) or because diagonality has genuinely increased the upper level lifetime (larger τ).

All designs in this generation featured two or more changes, thus none of the designs explored are well controlled with respect to any one variable. This, unfortunately, limits the conclusions that can be drawn from this generation.

4.2.5 Fourth generation designs

The emphasis of the fourth generation designs was on suppressing lower level parasitics, in a bid to correct the early NDR problem identified in the 3rd generation designs. At the same time, encouraged by results from 3rd generation designs demonstrating that low oscillator strength lasing is possible, these designs were made even more diagonal.

The low injection anticrossings of these designs have resulted in some of the lowest THz QCL threshold currents ever seen at MIT. Unfortunately, this generation of

devices almost certainly had growth problems, exhibiting many anomalies in transport data.

OWI215G-M4

OWI215-M4 is essential an a version of OWI222G with a smaller collector anticrossing (4.52meV down to 3.68 meV). Other major anticrossings and the oscillator strength are kept essentially unchanged, although the lasing frequency is slightly lower. Design parameters and experimental results are shown in figures 4-38 and 4-39 respectively.

Although the $1' - 2$ parasitic anticrossing only slightly lower than OWI222G (0.44meV versus 0.50meV), thickening the collector barrier appears to have significantly reduced subthreshold parasitics. Threshold and parasitic current densities are approximately half that of OWI222G. It had T_0=145K, which is similar to OWI222G. Unfortunately, $J_{\max}$ is reduced by more than half as well. As such, OWI215G-M4 only attained $T_{\max}$=126K.

While the IVs of OWI215G-M4 do not present any obvious abnormalities, GV measurements nevertheless suggest that there may have been some growth problems in this design. Examining figure 4-39d, the subthreshold region after the $1' - 2$ parasitic shows a number of small bumps in conductance. As this is a one-well injector design, there should be no significant parasitic level alignments following $1' - 2$, and therefore this suggest some inhomogeneities in bias.

OWI220G-M5

OWI220G-M5 is ostensibly a 20% aluminum version of OWI222G, but there are significant differences in design parameters. There were two growths of OWI220G-M5: VB0412 was undergrown by 1.25%. amd VB0409 was overgrown by 2.6%. Their experimental results are very similar, so only results from VB0412 are presented here. Design specifications and experimental results for OWI220G-M5 are shown in figures 4-40 and 4-41.

Unfortunately, growth for both VB0409 and VB0412 is suspect. Their IVs show multiple NDR behavior. In CW, IVs show "saw-tooth" behavior. This is attributed

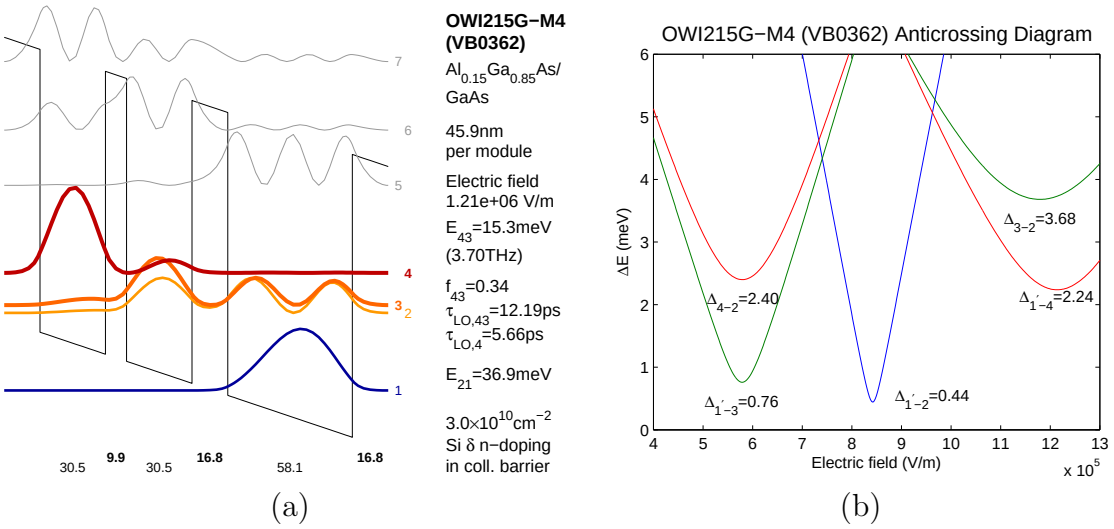


Figure 4-38: Design parameters for OWI215G-M4, wafer VB0362 (undergrown by 1.47%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Injection anticrossing is $\Delta_{1'-4}$. Collector anticrossing is Δ_{3-2}.

to drift in growth rate, based on results from older devices that precede this thesis.

Nevertheless, this design is promising. Figure 4-40d shows that the parasitic $1'-2$ anticrossing is smaller than OWI222G (0.34meV compared to 0.5meV), and indeed, the parasite knee in the IV is much less pronounced than in OWI222G. Moreover, despite the difficulties with growth, this design had the best temperature performance of devices in this generation, with $T_{\max}$=154K.

OWI210H

OWI210H is a very diagonal version of OWI222G. Two wafers were grown: VB0373 (undergrown by 2.3%) and VB0364 (undergrown by 6.4%). Compared to OWI222G, VB0373 has an oscillator strength reduced from 0.37 to 0.20. It also has a reduced injection anticrossing (from 2.15meV down to 1.64 meV) and collector anticrossing (from 4.52meV down to 4.00meV). VB0364 also has a reduced oscillator strength, 0.24. After accounting for VB0364's extreme undergrowth, VB0364 actually has approximately the same injector anticrossing (2.02meV). The collector anticrossing is slightly larger. Design specifications are shown in figures 4-42. Experimental results for VB0373 and VB0364 are shown in figures 4-43 and 4-44 respectively.

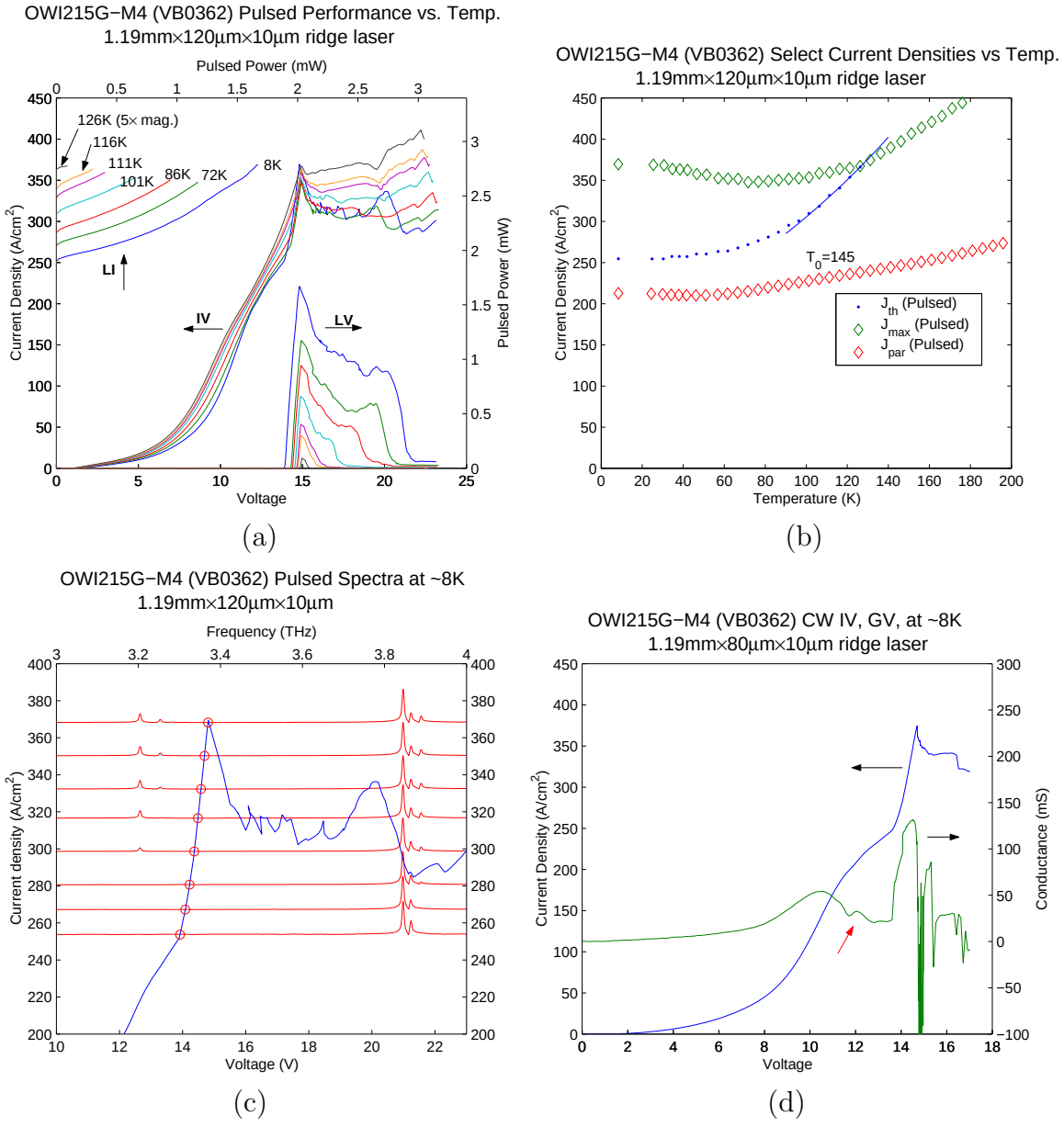
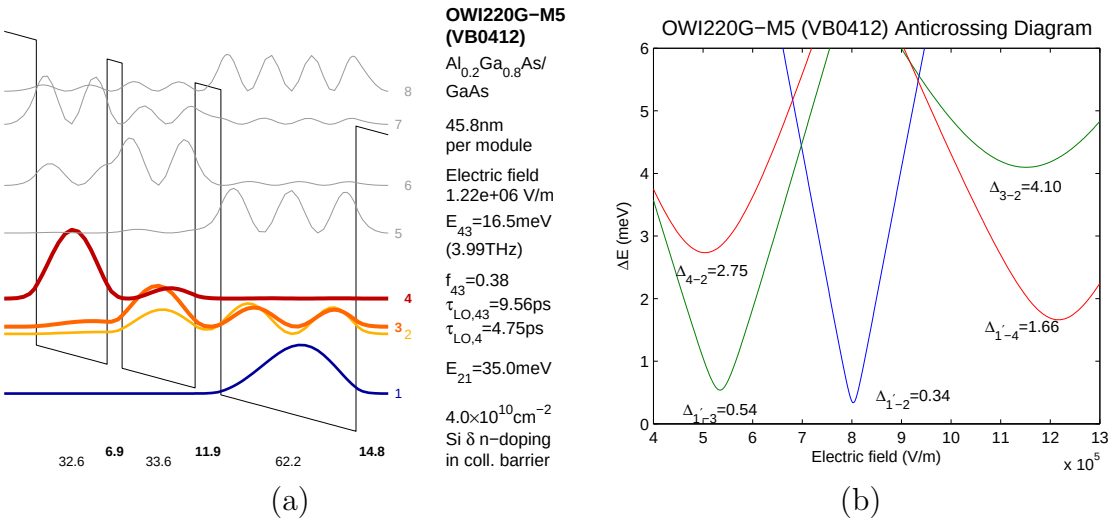


Figure 4-39: Experimental results for OWI215G-M4, wafer VB0362 (undergrown by 1.47%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 12.5V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) CW GIV at ∼8K. Red arrow indicates location of abnormality in GV.



(a) (b)

Figure 4-40: Design parameters for OWI220G-M5, wafer VB0412 (undergrown by 1.25%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-3}$. Collector anticrossing is Δ_{4-3}. (c) Postulated band diagram based on experimental lasing frequency. From (b), one sees that this bias is closer to the optimal collection bias rather than the injection bias.

VB0373's experimental results are well behaved, so they are discussed first. Parasitics are well suppressed, enabling extremely low threshold current densities ($\sim$ 125A/cm^2). Interesting, $J_{\max}$ for this design is not that much reduced compared to OWI215G-M4. OWI210H, VB0373, differs from OWI215G-M4 primarily in having lower oscillator strength and injector anticrossing, both of which ought to reduce $J_{\max}$. The only other notable difference is a higher collector anticrossing; if this is the what causes the surprisingly high $J_{\max}$ for OWI210H, then it again suggests that resonant tunneling collection is a limiting step in transport.

OWI210H, VB0373, also exhibited a significantly lower T_0 than OWI215G-M4. While acknowledging that T_0 can vary widely even amongst devices from the same fabrication, and that VB0373 was measured only once, this is still surprising. This contradicts the expectation that increased diagonality should result in slower rise in threshold, but on the other hand, OWI210H achieved a slightly higher $T_{\max}$ than OWI215G-M4 (131K versus 126K) despite the smaller T_0. Finally, GV data again presents a slight abnormality in the subthreshold region.

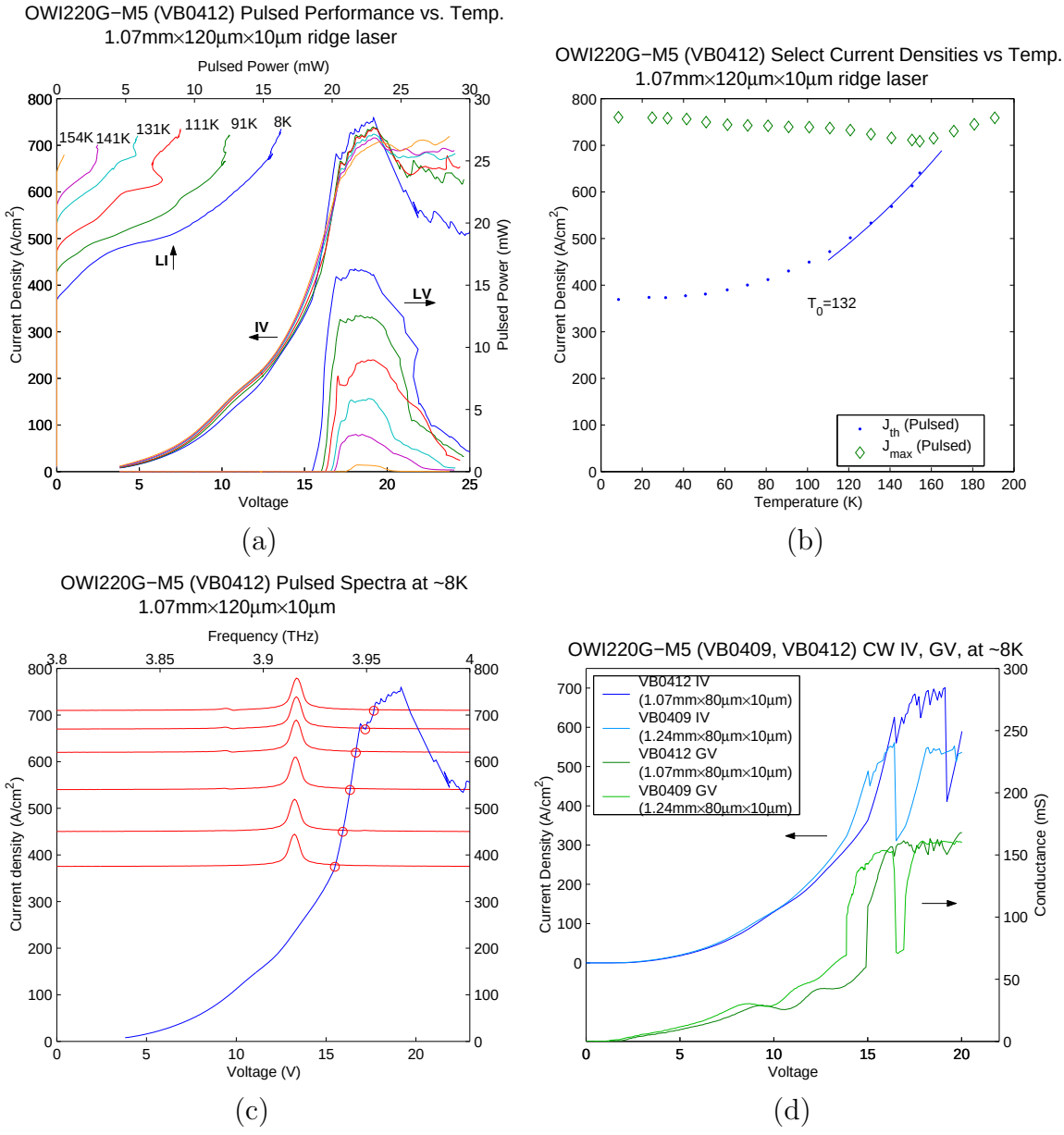


Figure 4-41: Experimental results for OWI220G-M5, wafer VB0412 (overgrown by 1.47%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold ($J_{\rm th}$) and maximum ($J_{\rm max}$) current densities versus temperature. There is no obvious parasitic bias. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicated bias location. (d) CW GIV at ∼8K for OWI220G-M5, VB0409 and VB0412, exhibiting distinctive saw-tooth shape in IV and GV.

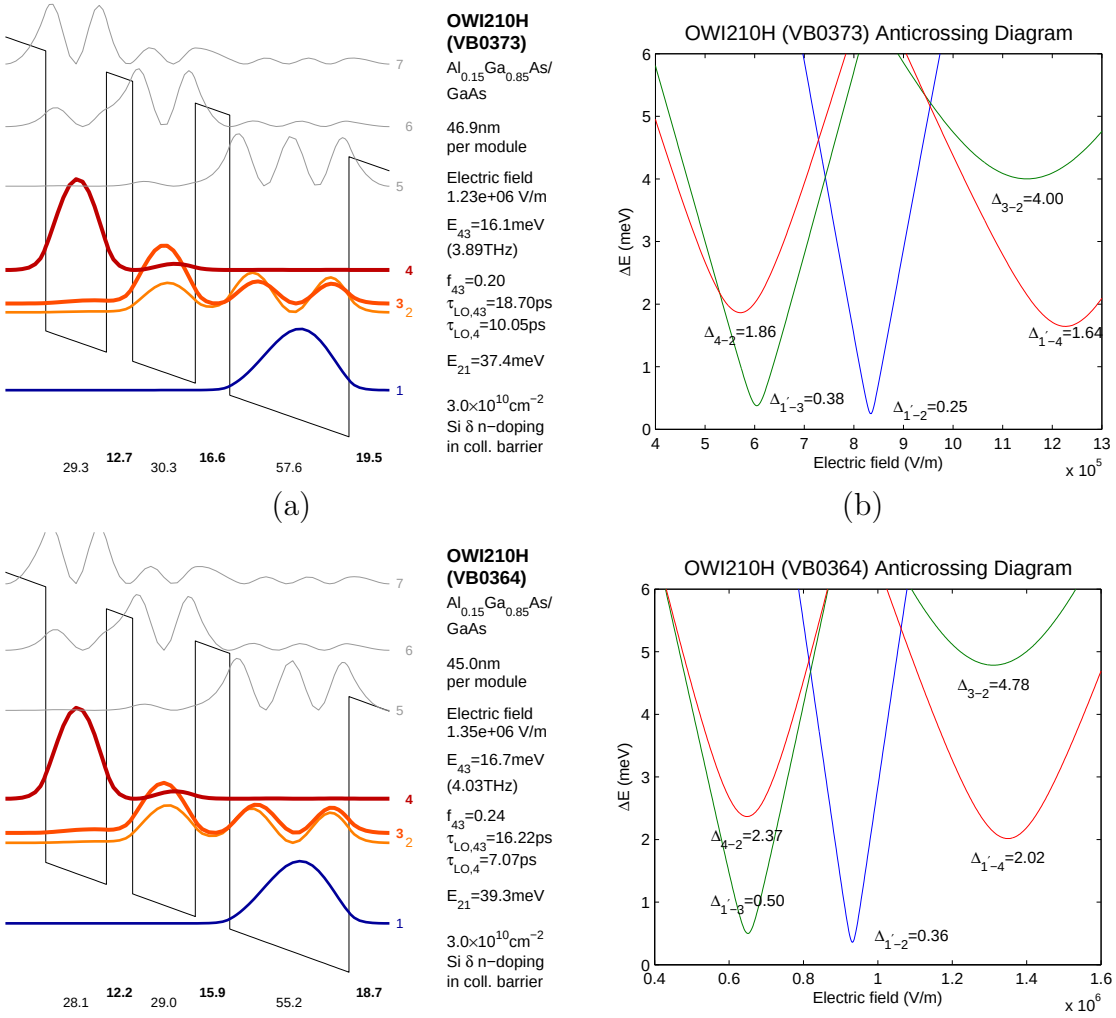


Figure 4-42: Design parameters for OWI210H, wafer VB0373 (undergrown by 2.3%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}. Injection anticrossing is $\Delta_{1'-4}$. (c) and (d) are the same plots for VB0364.

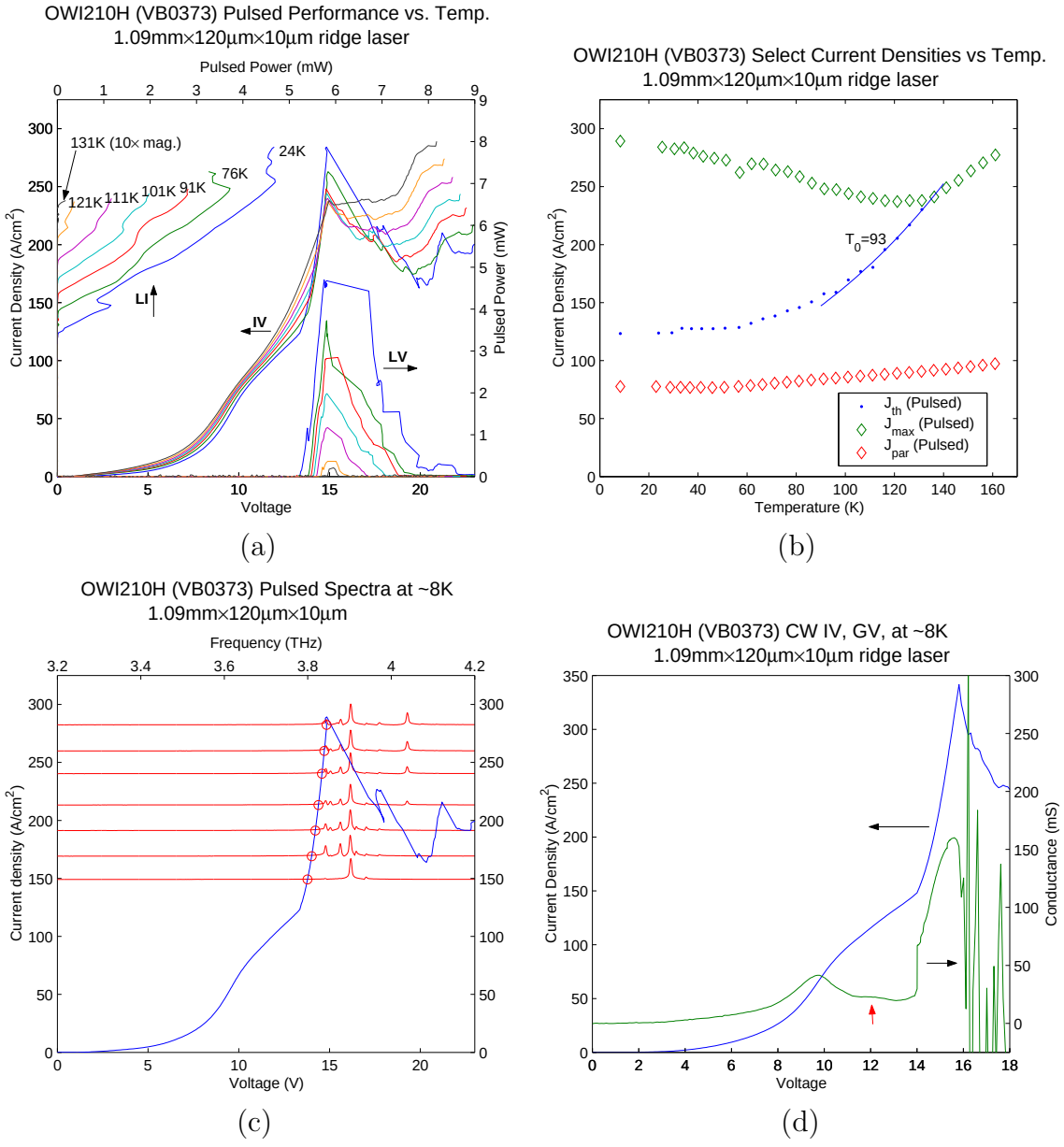


Figure 4-43: Experimental results for OWI210H, wafer VB0373 (undergrown by 2.3%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from IVs at 10.5V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) CW GIV at ∼8K, red arrow indicates (slight) abnormality in GV.

In contrast, VB0364 had extremely unusual experimental results. Lasing occurred at extremely high biases, deep in NDR. However, surprisingly, all currents are approximately the same as for VB0373, which is inconsistent with the larger anticrossings predicted in the band diagram. Yet the high temperature (say, 131K) IV looks similar to that of VB0373, with an NDR occurring at ∼15V. Referring to figure 4-44c, however, the lasing frequency is higher than VB0373, as expected from the band diagram. Unfortunately, GV data is not available for this device, but nonetheless growth problems are suspected. A more detailed explanation for these abnormalities is not known.

OWI209H-M1

Continuing the same trend as OWI210H, OWI209H-M1 is even more diagonal and possesses an even lower injector anticrossing. Design specifications and experimental results are shown in figures 4-45 and 4-46 respectively.

The lower level parasitic knee is almost invisible in the IVs, enabling extremely low threshold current densities ($\sim 80\text{A/cm}^2$). Unfortunately, the wafer for OWI209H-M1 almost certainly had growth problems, with the majority of its lasing seemingly occurring in NDR, and possessing a very rough IV. CW measurements of GV and IV are also shown in figure 4-46d; these exhibited saw-tooth oscillations suggestive of growth inhomogeneity. It lased only up to $T_{\max}$=101K.

In its GV curve (figure 4-46d), OWI209H-M1 also had an exceptionally wide subthreshold bias region following its $1'-2$ alignment. This suggests that it possessed unusually low gain, consistent with growth inhomogeneity.

OWI208H-M2

OWI208H-M2 was intended to be a replica of OWI209H-M1 with double the doping. Of course, growth inaccuracies lead to differences, so one must nevertheless calculate the band diagrams separately. Two wafers of OWI208H-M2 were processed: VB0393 (undergrown by 3.2%), and VB0385 (overgrown by 4.25%). The hope was that doping would increase gain more than it would increase loss. The design specifications and

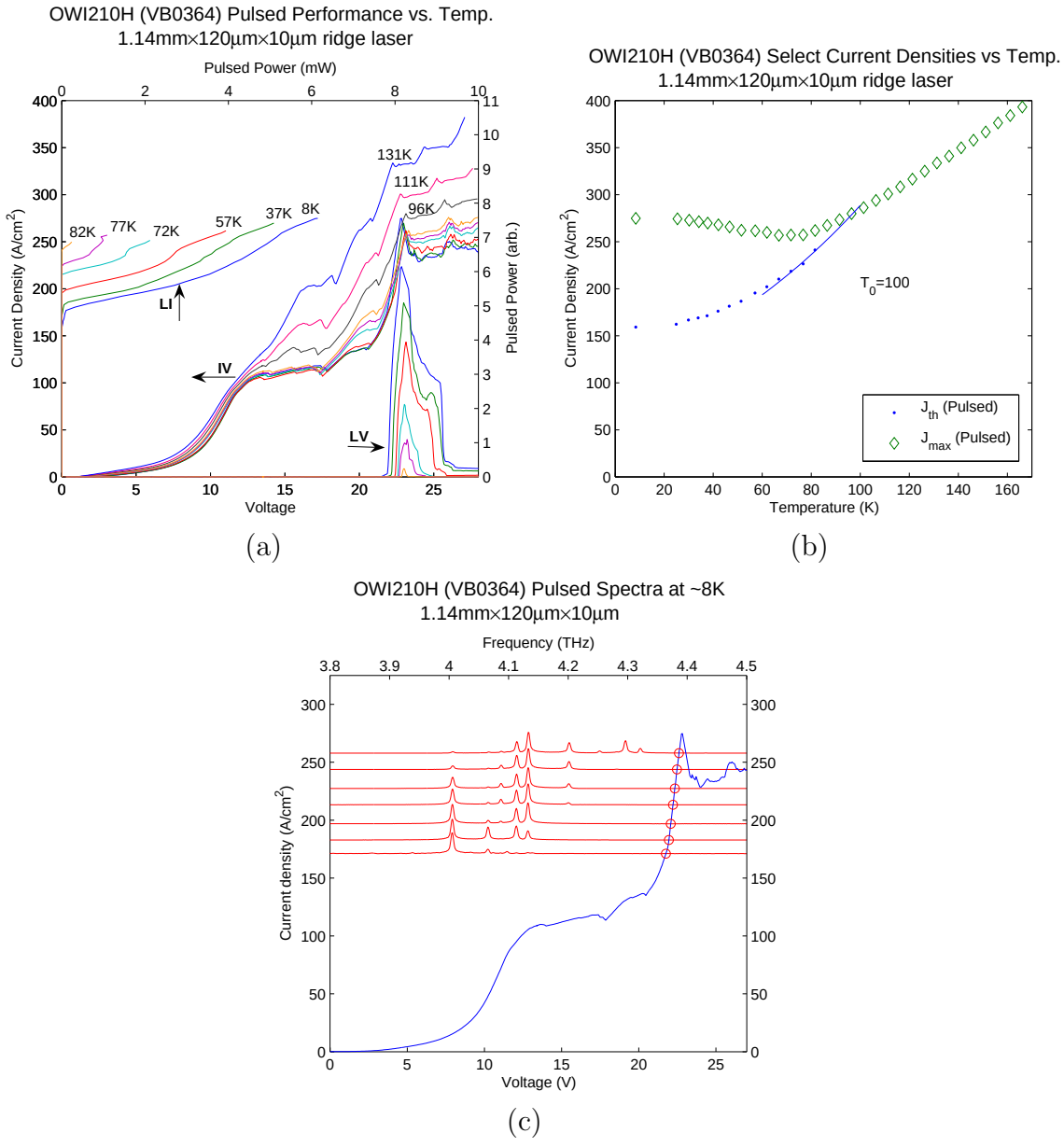
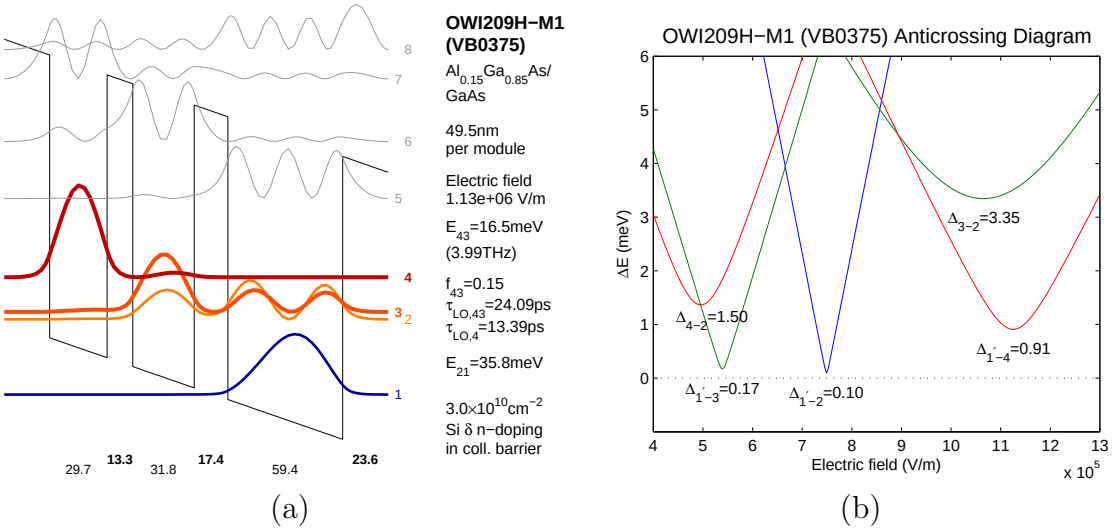


Figure 4-44: Experimental results for OWI210H, wafer VB0364. VB0364 is ostensibly undergrown by 6.4%, but is suspected to have other growth problems. (a) Pulsed LIV versus temperature measured for a wide device. Three post-lasing IVs are shown. (b) Threshold (J_{th}) and maximum (J_{max}) current densities versus temperature. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location.



(a) (b)

Figure 4-45: Design parameters for OWI209H-M1, wafer VB0375 (overgrown by 2.44%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-3}$. Collector anticrossing is Δ_{3-2}. Injector anticrossing is $\Delta_{1'-4}$.

experimental data for VB0393 are presented in figures 4-47 and 4-48 respectively. Figures 4-49 and 4-48 present the same for VB0385.

Contrary to the expectation that doubling the doping should essentially double the current, VB0393 exhibited approximately seven-fold higher currents than OWI209H-M1, VB0375. However, this result becomes reasonable when one considers that OWI209H-M1, VB0375, was overgrown, while OWI208H-M2, VB0393, was undergrown. Taken together, the layer width sequence differs significantly between the two wafers. This is reflected in the larger anticrossings of OWI208H-M2, VB0393 (injector anticrossing of 1.23meV versus 0.91meV in OWI209H-M2; collector anticrossing of 4.27meV versus 3.35meV in OWI209H-M2).

Growth problems are suspected in VB0393, on account of its irregular IVs. CW GIV measurements present in figure 4-48d also hint at possibility of the saw-tooth oscillations seen in other devices in this generation. Unfortunately, the high currents of this wafer invariably destroyed devices before data at higher biases could be acquired to confirm the saw-tooth oscillations. The GV also possessed two peaks in the subthreshold region, which is unexpected for an OWI design. Nonetheless, it achieved $T_{\max}$=125K, which is fairly high in comparison to other devices in this generation.

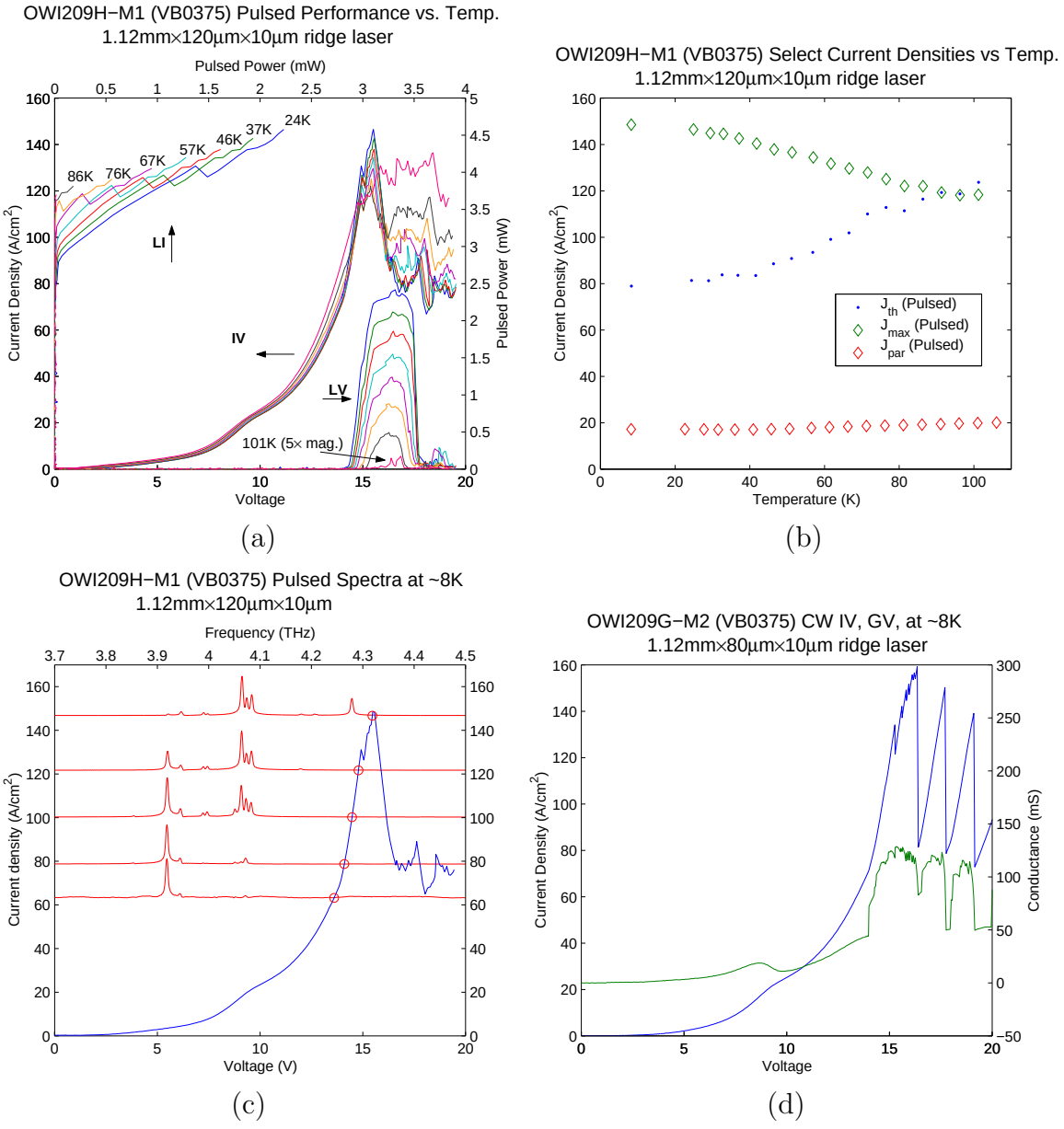
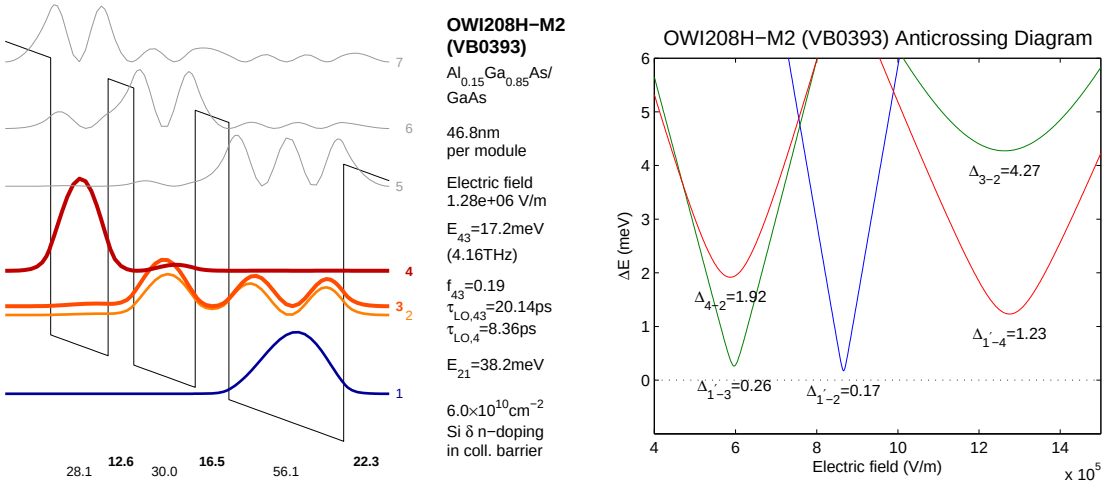


Figure 4-46: Experimental results for OWI209H-M1, wafer VB0375 (overgrown by 2.44%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}), maximum (J_{max}) and parasitic (J_{par}) current densities versus temperature. J_{par} is evaluated from the IVs at 9.6V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) CW GIV at ∼8K, showing distinctive saw-tooth pattern.



(a) (b)

Figure 4-47: Design parameters for OWI208H-M2, wafer VB0393 (undergrown by 3.2%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}. Injector anticrossing $\Delta_{1\prime-4}$.

OWI208H-M2, VB0385, renders a better comparison to OWI209H-M1, VB0375, as VB0385 was also overgrown. Comparing figures 4-49b and 4-45b, one sees that all anticrossings are similar except for slight decreases in the collect and injector anticrossings. In this case, the currents for VB0385 shown in figure 4-50a are ∼2.5 times that of VB0375 shown in figure 4-46a, approximately consistent with the doubled doping. However, the dynamic range of OWI208H-M2, VB0385 is much reduced compared to OWI209H-M2, VB0375 ($J_{\rm th}$ is ∼4 times that of VB0375). Indeed, only $T_{\rm max}$=51K was achieved. This argues against doubling the doping in future devices.

If one accepts that doubling the doping harmed performance, then it is remarkable that OWI208H-M2, VB0393, nevertheless outperformed OWI209H-M1. One possibility is that OWI208H-M2's superior performance was enabled by its better current transport, suggesting that injector anticrossings in this generation may have been too aggressively reduced; admittedly, however, many parameters are different between the two designs, so no definite conclusions can be made.

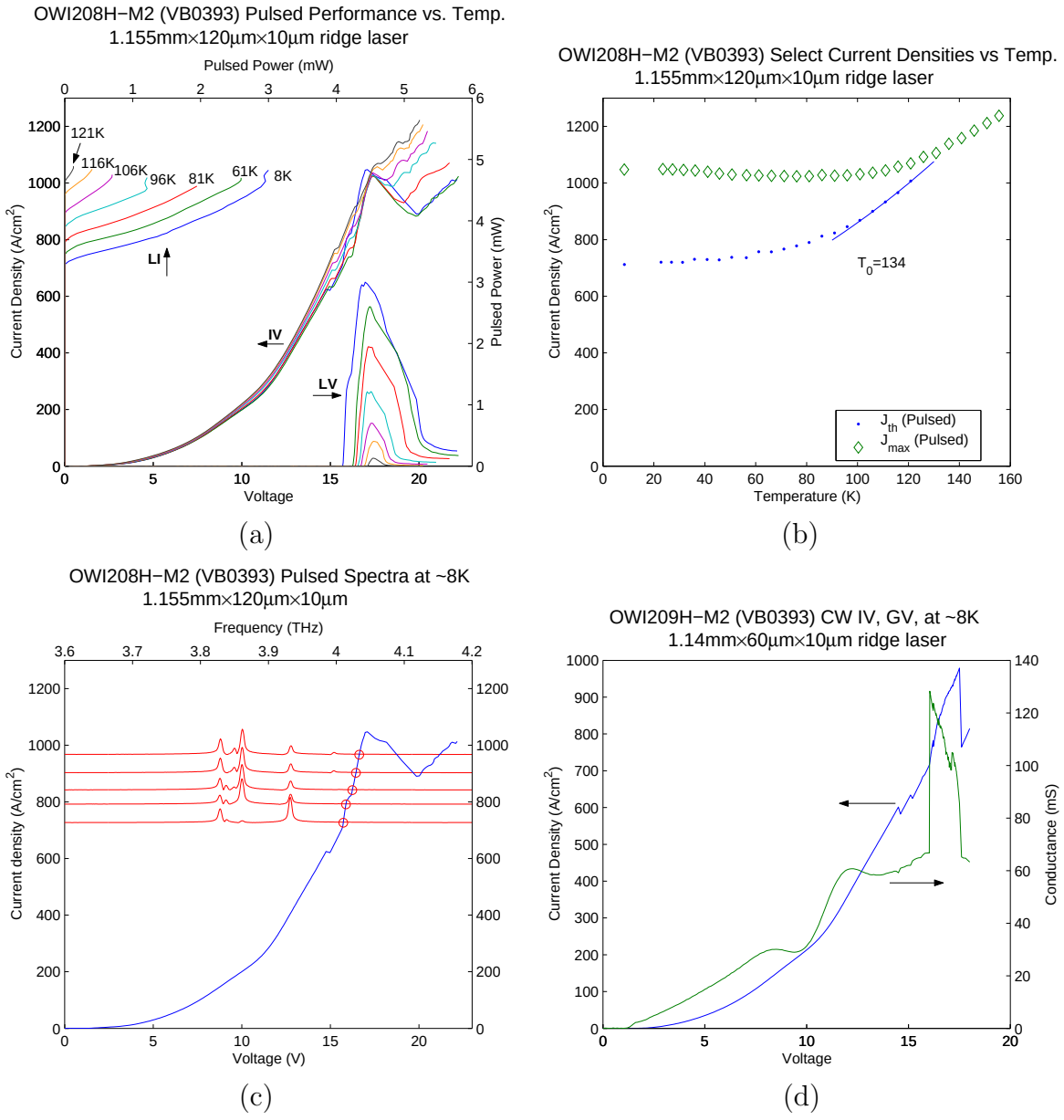
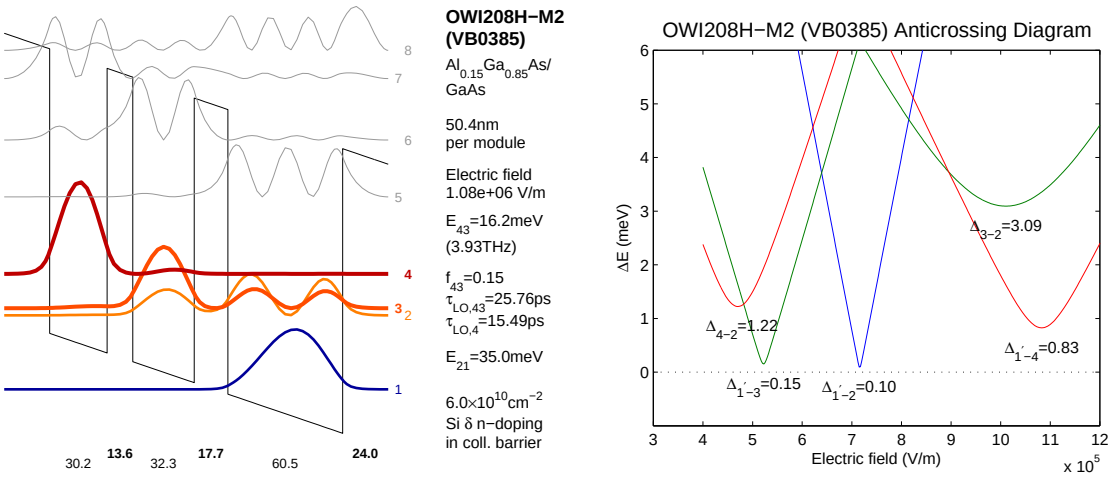


Figure 4-48: Experimental results for OWI208H-M2, wafer VB0393 (undergrown by 3.2%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}) and maximum (J_{max}) current densities versus temperature. There is no apparent parasitic bias. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicated bias location. (d) CW GIV at ∼8K, showing signs of saw-tooth pattern. GIV at higher biases could not be measured as the high currents typically destroyed the device.



(a) (b)

Figure 4-49: Design parameters for OWI208H-M2, wafer VB0385 (overgrown by 4.25%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Main parasitic in IV is attributed to $\Delta_{1'-2}$. Collector anticrossing is Δ_{3-2}. Injector anticrossing is $\Delta_{1'-4}$.

SDIP2W

SDIP2W is a descendant of TW246 designed with 20% aluminum barriers (SD stands for "super-diagonal"). Compared to TW246, all anticrossings are smaller. The injector anticrossing was reduced from 1.67meV to 1.21meV. The lower level parasitics anticrossings $\Delta_{1'-2}$ and $\Delta_{3'-4}$ were reduced from 0.53meV and 0.62meV to 0.26meV and 0.14meV, respectively. Design specifications and experimental results are shown in figures 4-51 and 4-52 respectively.

In principle, this design should have performed better. Unlike TW246 and TW260-M1, the injection anticrossing is fairly well separated in bias from parasitic anticrossings. The higher barriers should have reduced leakage to continuum in any event. However, this design performed poorly. The device whose results are shown in figure 4-52 lased only up to 76K (another device lased up to 96K, but its results are not shown because severe abnormalities in the data suggest that errors may have occurred during data collection). $J_{\max}$ for this device exhibits the drop versus temperature characteristic of most highly diagonal designs. Unlike TW246, the bias point of peak optical power does not lag behind the bias point of $J_{\max}$ with increasing tem-

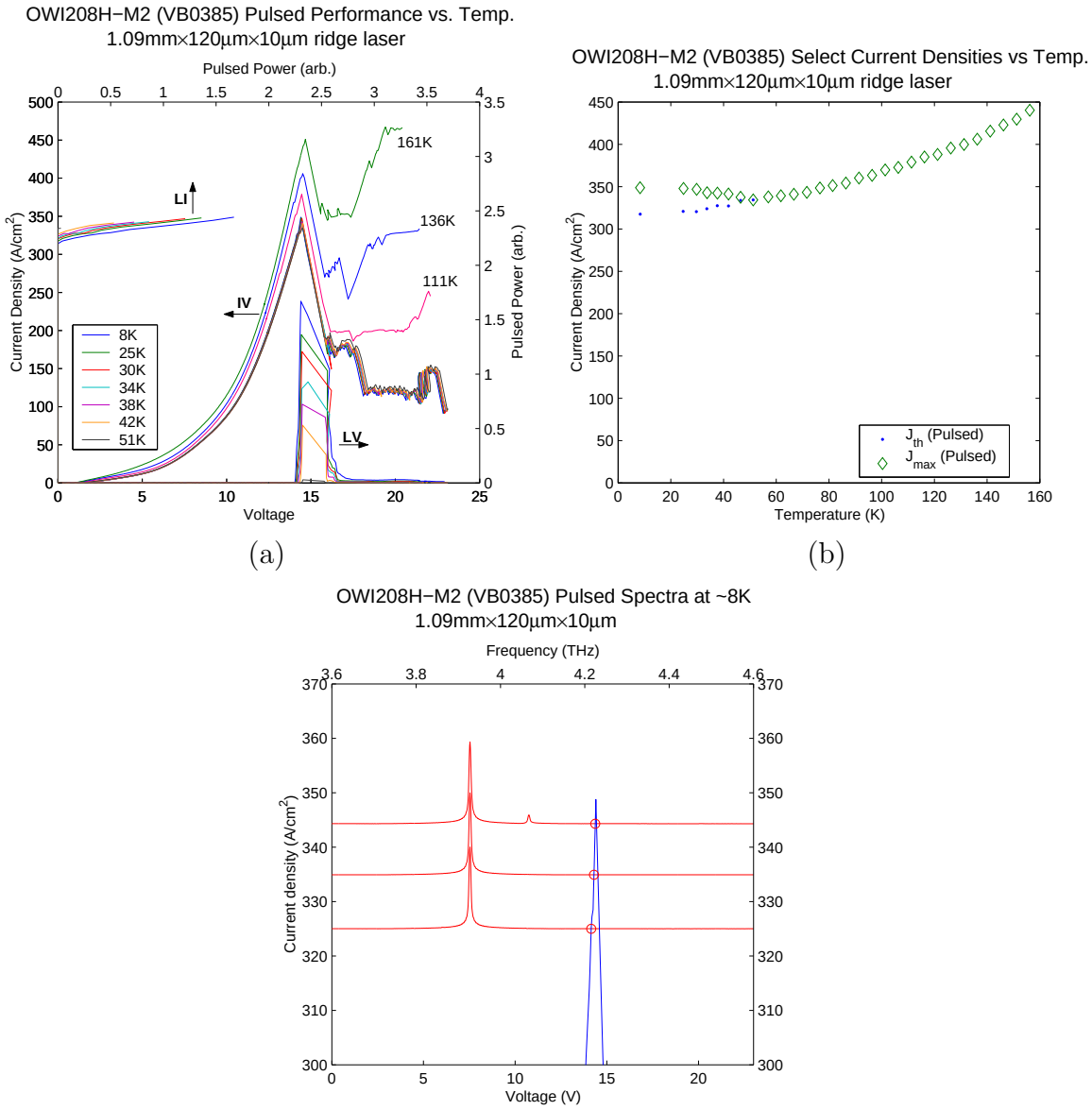
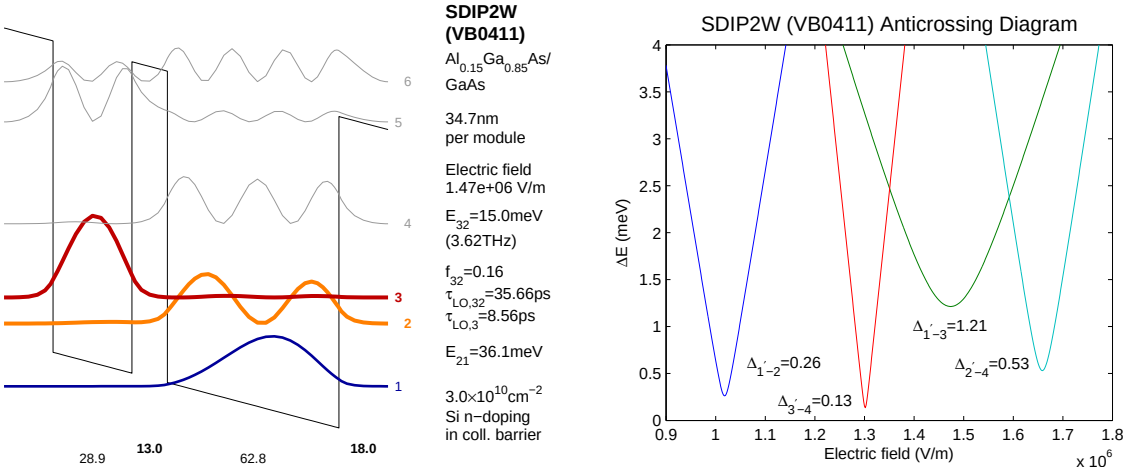


Figure 4-50: Experimental results for OWI208H-M2, wafer VB0385 (overgrown by 4.25%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold (J_{th}) and maximum (J_{max}) versus temperature. There is no apparent parasitic bias. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location.

perature (in TW246 this was cited as evidence for hot phonon absorption, or at least behavior consistent with hot phonon absorption).

Furthermore, SDIP2W lased at higher than expected frequencies. *GV* data in figure 4-52d also reveals a large bump in the subthreshold region. If this bump corresponds to injector anticrossing, then the device is lasing past its design bias. This is consistent with the unexpectedly high lasing frequency, but raises the question why there is no NDR corresponding to the injector detuning? On the other hand, the subthreshold bump could be the $3' - 4$ anticrossing.



(a) (b)

Figure 4-51: Design parameters for SDIP2W, wafer VB0411 (undergrown by 0.26%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Injector anticrossing is $\Delta_{1'-3}$. As this is a DSL design, there is no collector anticrossing.

SDRP4W

SDRP4W is essentially a very diagonal version of the classical FL family resonant phonon designs. However, all of its major parameters are changed with respect to previous FL designs, so a differential analysis with reference to older designs is difficult. Design specifications and experimental data are shown in figures 4-53 and 4-54 respectively.

The best feature of this design is its excellent suppression of parasitics; in fact, from figure 4-53b, the important $1' - 3$ anticrossing is approximately zero, and the

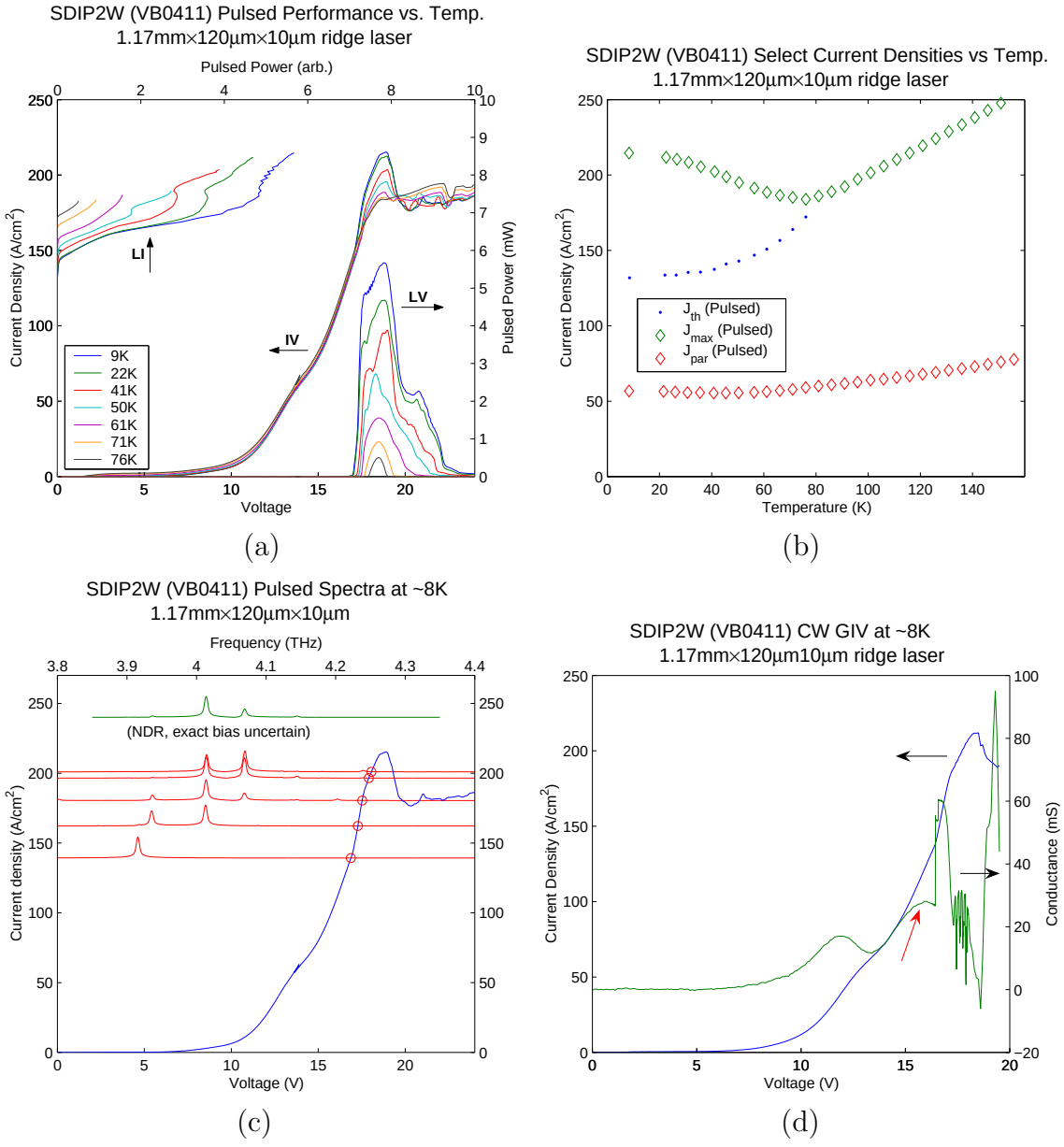
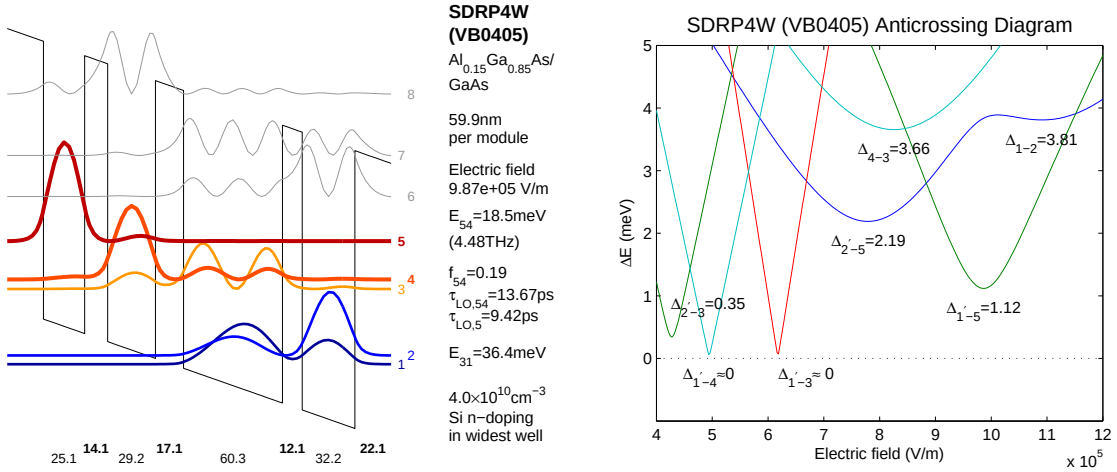


Figure 4-52: Experimental results for SDIP2W, wafer VB0411 (undergrown by 0.26%). (a) Pulsed LIV versus temperature measured for a wide device. (b) Threshold ($J_{\rm th}$), maximum ($J_{\rm max}$) and parasitic ($J_{\rm par}$) current densities versus temperature. $J_{\rm par}$ is evaluated from the IVs at 13.5V. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) CW GIV at ∼8K. Red arrow indicates location of unexpected bump before threshold.

$2'-3$ anticrossing is also small in comparison to, say, FL190S-M2. Its low temperature threshold current density is similar to that of FL190S-M2. Also, it exhibited FL190S-M2's rise in current versus temperature across all device biases.

Unfortunately, SDRP4W almost certainly suffered from growth problems, possessing a very rough IV, a large amount of NDR lasing, and saw-tooth oscillations in CW GIV measurements. It attained only $T_{\max} \approx$47K. Also, lasing occurs at much higher frequencies than expected. Better agreement is found by calculating the band diagram at higher bias. In figure 4-53b, the next logical bias is that corresponding to 1-2 anticrossing. But this still yields frequencies too low, so the results are not displayed.

The anticrossing diagram in figure 4-53b shows that the $1'-5$ injection anticrossing is badly misaligned with $2'-5$ injection anticrossing and the $4-3$ injector anticrossing. However, this feature is intentional; this "upward diagonal" injector scheme was successfully used by Williams to improve the performance of older devices [46].



(a) (b)

Figure 4-53: Design parameters for SDRP4W, wafer VB0405 (overgrown by 0.53%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Collector anticrossing is Δ_{4-3}. Injector anticrossings are $\Delta_{2'-5}$ and $\Delta_{1'-5}$.

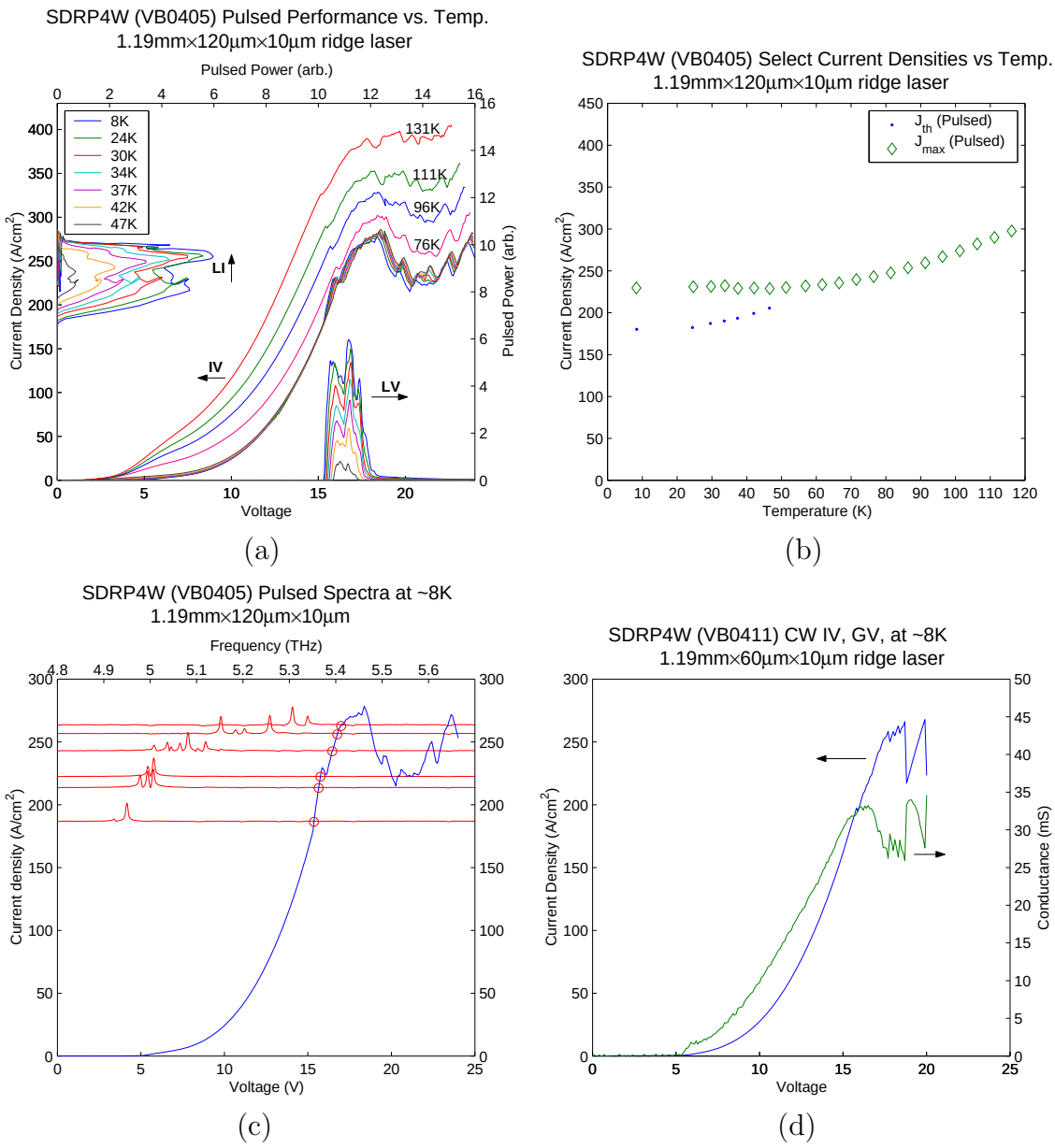


Figure 4-54: Experimental results for SDRP4W, wafer VB0405 (overgrown by 0.53%). (a) Pulsed LIV versus temperature measured for a wide device. A few non-lasing IVs at high temperatures are also displayed to demonstrate the rise in currents with temperature (these are marked outside of the legend box). (b) Threshold (J_{th}) and maximum (J_{max}) and current densities versus temperature. There is no obvious parasitic bias. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) CW GIV at ∼8K, showing distinctive saw-tooth pattern.

SDRP5W

SDRP5W is a resonant phonon design with 3 injector levels. The motivation for using a multi-subband injector is, again, to cut down on parasitics. Design specifications and experimental results are shown in figures 4-55 and 4-56 respectively.

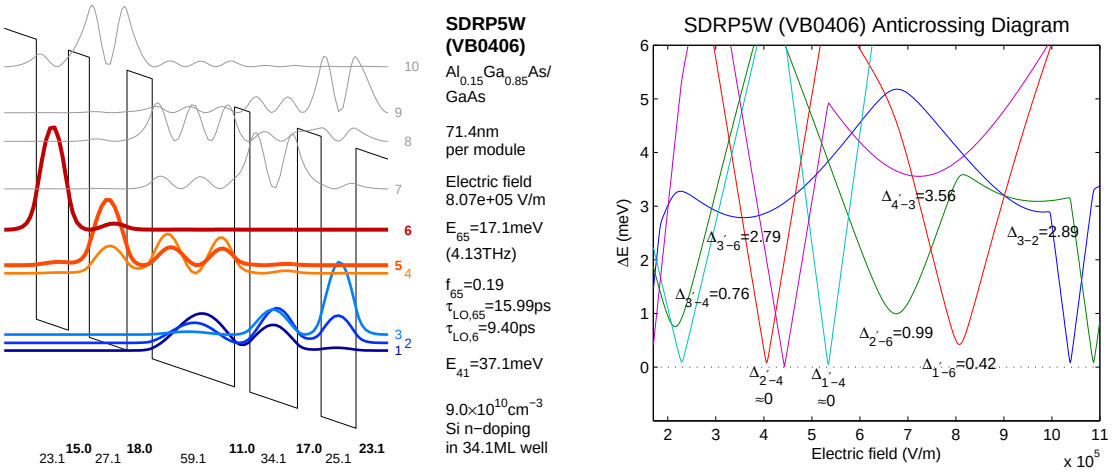
SDRP5W is promising in that it is the most successful design in this generation at cutting down subthreshold parasitics. Its threshold current density is the lowest of any MIT design, at $\sim 50\text{A/cm}^2$. Unfortunately, it again had inferior temperature performance, with $T_{\max}$=96K. Figure 4-56a shows that currents rise steeply with temperature; indeed, T_0 is only 36K for this design. The reasons for this increase is not well understood, but one possibility is that thermally activated population of level 3 leads to increased prominence of the $3'-4$ parasitic channel at higher temperatures. The $3'-4$ anticrossing is $\Delta_{3'-4} = 0.76\text{meV}$, whereas the $1'-4$ and $2'-4$ anticrossings have negligible anticrossing gaps.

SDRP5W also has unusual transport characteristics. Although it lacked a distinct NDR, CW *GIV* measurements reveal extremely irregular conductance post-threshold. In view of the other wafers grown in this generation, there is a possibility that SDRP5W suffers from growth problems as well, although the evidence is less definite than in the OWI designs previously discussed.

SDRP5W also lased at higher than expected frequencies. A better match to experimental lasing frequencies is obtained with the band-diagram calculated at a higher bias, close to the injector level anticrossing (3-2 anticrossing in figure 4-55b). But this is strange, as at such high biases neither collection nor injection is anticrossed.

Summary

Fourth generation devices generally perform even worse than third generations devices, and moreover are plagued by issues of growth. Nevertheless, they succeed in cutting down subthreshold parasitics and preventing the early NDR seen in third generation designs. This has resulted in some of the lowest threshold current densities ever witnessed in THz QCLs at MIT, but these are unfortunately accompanied by



(a) (b)

Figure 4-55: Design parameters for SDRP5W, wafer VB0406 (overgrown by 0.23%). (a) Calculated single-module band diagram at optimal injection bias. (b) Calculated anticrossings. Collector anticrossing is Δ_{5-4}, and injector anticrossing is $\Delta_{1'-6}$.

large reductions in maximum current flow as well. Whereas third generation devices are hypothesized to be limited by the early onset of NDR, fourth generation devices appear to be limited by stunted transport. It may be desirable to take some middle ground between third and fourth generation devices and increase injector anticrossings somewhat, or to find other methods of suppressing parasitics.

Furthermore, $J_{\max}$ continues to drop with temperature for all devices during lasing, despite the extremely small injection anticrossings of this generation. This suggests that diagonality has genuinely increased upper laser level lifetime, but it remains a mystery why threshold rises so rapidly with temperature.

Other findings include promising results for a 20% aluminum barrier variant of OWI222G, and experimental evidence suggesting that doubling standard doping levels harms performance.

4.3 Conclusions

The major results of this thesis are as follows.

- Highly diagonal THz quantum cascade lasers have been demonstrated to be

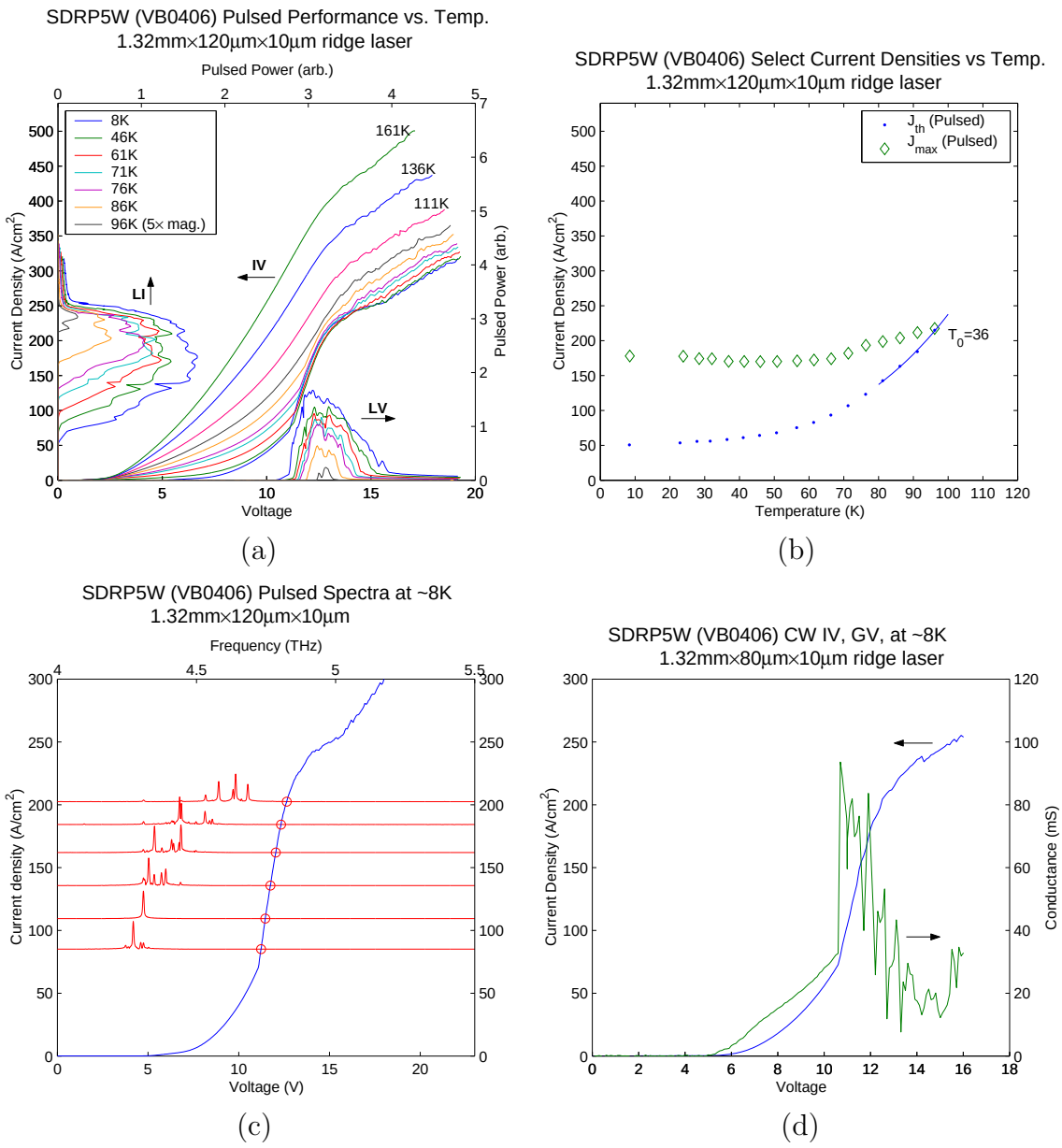


Figure 4-56: Experimental results for SDRP5W, wafer VB0406 (overgrown by 0.23%). (a) Pulsed LIV versus temperature measured for a wide device. Several non-lasing IVs are shown to illustrate the rise in current with temperature. (b) Threshold (J_{th}), maximum (J_{max}) current densities versus temperature. Because there is no clear NDR in the IV, J_{max} is chosen to correspond to peak optical power. There is no obvious parasitic bias. (c) Pulsed spectra at ∼8K, overlaid on corresponding IV to indicate bias location. (d) CW GIV at ∼8K, showing roughness post-threshold.

capable of lasing—a hollow victory perhaps, since said devices performed poorly in almost every instance. But the designs in this thesis were never well controlled with respect to individual design parameters, so diagonality as a path to high temperature performance cannot be conclusively ruled out. Cause for optimism remains.

- LO phonon scattering assisted (SA) injection in THz quantum cascade lasers has been demonstrated for the first time. The SA injection based lasing identified in OWI185E-M1 has soundly broken the empirical $T_{\max} < \hbar\omega/k_B$ limit, suggesting that SA injection may prove superior to conventional resonant tunneling injection.

- Although not definitive, all available evidence suggests that thermal backfilling is irrelevant at present operating temperatures, suggesting that design efforts be focused elsewhere.

The major recommendation for future designs is tweaking fewer design parameters. A continual frustration in chapter 4 is the difficulty of interpreting experimental results in light of the aggressive changes made between device generations. Hopefully, a more conservative design strategy will enable realization of the theoretical benefits of diagonality, or at least definitively determine if high diagonality is an inviable route to high temperature performance.

Bibliography

[1] M. Tacke. Lead-salt lasers. *Phil. Trans. R. Soc. Lond. A*, 359:547, 2001.

[2] The terahertz gap: the generation of far-infrared radiation and its applications. *Phil. Trans. R. Soc. Lond. A*, 362(1815), 2004. Special issue with a compilation of 16 papers.

[3] Photonic terahertz technology. *Semicond. Sci. Technol.*, 20(7), 2005. Special issue with a compilation of 22 papers.

[4] Terahertz for military and security application IV. *Proc. SPIE*, 6212, 2006. Special issue with a compilation of 33 papers.

[5] P. H. Siegel. Terahertz technology. *IEEE Trans. Microwave Theory Tech.*, 50:910, 2002.

[6] P. H. Siegel. Terahertz technology in biology and medicine. *IEEE Trans. Microwave Theory Tech.*, 52:2438, 2004.

[7] I. Mehdi. THz local oscillator technology. *Proc. SPIE*, 5498:103, 2004.

[8] J. Ward, E. Schlecht, G. Chattopadhyay, A. Maestrini, J. Gill, F. Maiwald, H. Javadi, and I. Mehdi. Capability of THz sources based on Schottky diode frequency multiplier chains. *IEEE MTT-S International Microwave Symposium Digest*, 3:1587–111, 2004.

[9] H.-W. Hübers, S. G. Pavlov, and V. N. Shastin. Terahertz lasers based on germanium and silicon. *Semicond. Sci. Technol.*, 20:S221, 2005. Also see the cited references.

[10] M. Inguscio, G. Moruzzi, K. M. Evenson, and D. A. Jennings. A review of frequency measurements of optically pumped lasers from 0.1 to 8 THz. *J. Appl. Phys.*, 60:R161, 1986.

[11] L. Esaki and R. Tsu. Superlattice and negative differential conductivity in semiconductors. *IBM J. Res. Dev.*, 14:61, 1970.

[12] R. F. Kazarinov and R. A. Suris. Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice. *Sov. Phys. Semicond.*, 5:707, 1971.

[13] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho. Quantum cascade laser. *Science*, 264:553, 1994.

[14] J. S. Yu, S. Slivken, A. Evans, S. R. Darvish, J. Nguyen, and M. Razeghi. High-power $\lambda \sim 9.5$ μm quantum-cascade lasers operating above room temperature in continuous-wave mode. *Appl. Phys. Lett.*, 88:091113, 2006.

[15] F. Capasso, R. Paiella, R. Martini, R. Colombelli, C. Gmachl, T. L. Myers, M. S. Taubman, R. M. Williams, C. G. Bethea, K. Unterrainer, H. Y. Hwang, D. L. Sivco, A. Y. Cho, A. M. Sergent, H. C. Liu, and E. A. Whittaker. Quantum cascade lasers: ultrahigh-speed operation, optical wireless communication, narrow linewidth, and far-infrared emission. *IEEE J. Quantum Electron.*, 38:511, 2002.

[16] R. Köhler, 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*, 417:156, 2002.

[17] G. Scalari, C. Walther, M. Fischer, R. Terazzi, H. Beere, D. Ritchie, and J. Faist. Thz and sub-thz quantum cascade lasers. *Laser & Photon. Rev.*, 3:45, 2009.

[18] A. Wade, G. Fedorov, D. Smirnov, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno. Magnetic-field-assisted terahertz quantum cascade laser operating up to 225 K. *Nature Photonics*, 3:41, 2008.

[19] M. Beck, D. Hofstetter, T. Aellen, J. Faist, U. Oesterle, M. Ilegems, E. Gini, and H. Melchior. Continuous wave operation of a mid-infrared semiconductor laser at room temperature. *Science*, 295:301, 2002.

[20] S. Kumar, Q. Hu, and J. L. Reno. 186k operation of terahertz quantum-cascade lasers based on a diagonal design. *Applied Physics Letters*, 94:131105, 2009.

[21] G. Scalari. *Magneto-spectroscopy and the development of terahertz quantum cascade lasers*. PhD dissertation, Université du Neuchâtel, Institut du Physique, 2005.

[22] C.-F. Hsu, P. Zory J.-S. O, and D. Botez. Intersubband quantum-box semiconductor lasers. *IEEE J. Sel. Top. Quant. Electron.*, 6:491, 2000.

[23] G. Sun, R. A. Soref, and J. B. Khurgin. Active region design of a terahertz GaN/$Al_{0.15}Ga_{0.85}N$ quantum cascade laser. *Superlattices Microstruct.*, 37:107, 2005.

[24] B. S. Williams. Terahertz quantum-cascade lasers. *Nature Photonics*, 1:517, 2007.

[25] S.-C. Lee and A. Wacker. Nonequilibrium green's function theory for transport and gain properties of quantum cascade structures. *Phys. Rev. B*, 66(24):245314, 2002.

[26] T. Kubis, C. Yeh, P. Vogl, A. Benz, G. Fasching, and C. Deutsch. Theory of nonequilibrium quantum transport and energy dissipation in terahertz quantum cascade lasers. *Phys. Rev. B*, 79(19):195323, 2009.

[27] J. Green, T. B. Boykin, C. D. Farmer, M. Garcia, C. N. Ironside, G. Klimeck, R. Lake, and C. R. Stanley. Quantum cascade laser gain medium modeling using a second-nearest-neighbor tight-binding model. *Superlattices and Microstructures*, 37(6):410, 2005.

[28] A. Wacker. Semiconductor superlattices: a model system for nonlinear transport. *Physics Reports*, 357(1):1, 2002.

[29] R. Terazzi, T. Gresch, M. Giovannini, N. Hoyler, N. Sekine, and J. Faist. Bloch gain in quantum cascade lasers. *Nature Physics*, 3:329, 2007.

[30] M. Glück, A. R. Kolovsky, and H. J. Korsch. Wannier-stark resonances in optical and semiconductor superlattices. *Physics Reports*, 366:103, 2002.

[31] N. Hatano, K. Sasada, and H. Nakamura. Some properties of the resonant state in quantum mechanics and its computation. *Progr. Theoret. Phys*, 119:187, 2008.

[32] S. Kumar. *Development of Terahertz Quantum Cascade Lasers*. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2007.

[33] H. Callebaut. *Analysis of the electron transport properties in quantum cascade lasers*. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2006.

[34] R. C. Iotti and F. Rossi. Nature of charge transport in quantum-cascade lasers. *Phys. Rev. Lett.*, 87:146603, 2001.

[35] H. Callebaut and Q. Hu. Importance of coherence for electron transport in terahertz quantum cascade lasers. *J. Appl. Phys.*, 98:104505, 2005.

[36] C. Weber, A. Wacker, and A. Knorr. Density-matrix theory of the optical dynamics and transport in quantum cascade structures: The role of coherence. *Phys. Rev. B*, 79:165322, 2009.

[37] A. Wacker. Coherence and spatial resolution of transport in quantum cascade lasers. *Physica Status Solidi (C)*, 5:215, 2008.

[38] P. Harrison. *Quantum wells, wires and dots*. John Wiley and Sons, Ltd, Chichester, UK, 1999.

[39] S. F.-P. Paul and H Fouckhardt. An improved shooting approach for solving the time-independent Schrödinger equation for III/V QW structures. *Phys. Lett. A*, 286:199, 2001.

[40] J. P. Boyd. *Chebyshev and Fourier Spectral Methods*. DOVER Publications, New York, 2nd edition, 2000.

[41] A. Patera. A spectral element method for fluid dynamics: Laminar flow in a channel expansion. *J. Comp. Phys*, 54:468, 1984.

[42] C. Cheng, Q. H. Liu, J.-H. Lee, and H. Z. Massoud. A spectral element method for the schrodinger poisson system. *J. Comp. Phys*, 3:417, 2004.

[43] J. A. Weideman and S. C. Reddy. A matlab differentiation matrix suite. *ACM Trans. Math. Softw.*, 26:465, 2000.

[44] M. McLennan and S. Datta. *SEQUAL User's Manual*. Purdue University, Layfayette, 1989.

[45] J. H. Smet, C. G. Fonstad, and Q. Hu. Intrawell and interwell intersubband transitions in multiple quantum wells for far-infrared sources. *J. Appl. Phys.*, 79:9305, 1996.

[46] B. S. Williams. *Terahertz quantum cascade lasers*. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2003.

[47] X. Gao. *Monte Carlo simulation of electron dynamics in quantum cascade lasers*. PhD dissertation, University of Wisconsin-Madison, Department of Electrical Engineering, 2008.

[48] J. H. Smet. *Intrawell and interwell intersubband transitions in single and multiple quantum well heterostructures*. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1995.

[49] C. Sirtori, F. Capasso, and J. Faist. Nonparabolicity and a sum rule associated with bound-to-bound and bound-to-continuum intersubband transitions in quantum wells. *Phys. Rev. B*, 50:8663, 1994.

[50] G. Scalari, L. Ajili, J. Faist, H. Beere, E. Linfield, D. Ritchie, and G. Davies. Far-infrared ($\lambda \cong 87\mu$m) bound-to-continuum quantum- cascade lasers operating up to 90 K. *Appl. Phys. Lett.*, 82:3165, 2003.

[51] 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. *Appl. Phys. Lett.*, 82:1015, 2003.

[52] B. Xu. *Development of intersubband terahertz lasers using multiple quantum well structures*. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1998.

[53] S. Kumar, B. S. Williams, , Q. Hu, and J. L. Reno. 1.9 THz quantum-cascade lasers with one-well injector. *Appl. Phys. Lett.*, 88:121123, 2006.

[54] M. A. Belkin, J. A. Fan, S. Hormoz, F. Capasso, S. P. Khanna, M. Lachab, A. G. Davies, and E. H. Linfield. Terahertz quantum cascade lasers with copper metal-metal waveguides operating up to 178 k. *Opt. Express*, 16(5):3242, 2008.

[55] R. Nelander and A. Wacker. Temperature dependence of the gain profile for terahertz quantum cascade lasers. *APL*, 92:081102, 2008.

[56] Y. Kawamura, K. Wakita, H. Asahi, and K. Kurumada. Observation of room temperature current oscillation in InGaAs/InAlAs MQW pin diodes. *Japanese Journal of Applied Physics*, 25(2):L928, 1986.

[57] H. Luo, S.R. Laframboise, Z.R. Wasilewski, and H.C. Liu. Effects of injector barrier on performance of terahertz quantum-cascade lasers. *Electronics Letters*, 43:633, 2007.

[58] H. C. Liu, M. Wächter, D. Ban, Z. R. Wasilewski, M. Buchanan, G. C. Aers, J. C. Cao, S. L. Feng, B. S. Williams, and Q. Hu. Effect of doping concentration on the performance of terahertz quantum-cascade lasers. *Appl. Phys. Lett.*, 87:111104, 2005.

[59] C. Sirtori, F. Capasso, J. Faist, A. L. Hutchinson, D. L. Sivco, and A. Y. Cho. Resonant tunneling in quantum cascade lasers. *IEEE J. Quantum Electron.*, 34:1722, 1998.

[60] S. Kumar and Q. Hu. Coherence of resonant-tunneling transport in terahertz quantum-cascade lasers. *Phys. Rev. B*, 80:245316, 2009.

[61] M. Yamanishi, K. Fujita, T. Edamura, and H. Kan. Indirect pump scheme for quantum cascade lasers: dynamics of electron-transport and very high t0-values. *Opt. Express*, 16:20748, 2008.

[62] H. Luo, S. R. Laframboise, Z. R. Wasilewski, G. C. Aers, H. C. Liu, and J. C. Cao. Terahertz quantum-cascade lasers based on a three-well active module. *Applied Physics Letters*, 90:041112, 2007.

[63] S. Kumar, C. W. I. Chan, Q. Hu, and J. L. Reno. Two-well terahertz quantum-cascade laser with direct intrawell-phonon depopulation. *Applied Physics Letters*, 95:141110, 2009.

[64] V. Spagnolo, M. S. Vitiello, G. Scamarcio, B. S. Williams, S. Kumar, Q. Hu, and J. L. Reno. Hot-phonon generation in thz quantum cascade lasers. *Journal of Physics: Conference Series*, 92(1):012018, 2007.

[65] R. C. Iotti, F. Rossi, M. S. Vitiello, G. Scamarcio, L. Mahler, and A. Tredicucci. Impact of nonequilibrium phonons on the electron dynamics in terahertz quantum cascade lasers. *Applied Physics Letters*, 97:033110, 2010.

[66] J. T. Lü and J. C. Cao. Monte carlo simulation of hot phonon effects in resonant-phonon-assisted terahertz quantum-cascade lasers. *Appl. Phys. Lett.*, 88:061119, 2006.