Analysis of the Electron Transport Properties in Quantum Cascade Lasers

by

Hans Callebaut

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

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

April 2006
[June 2006]

Author ....
Department of Electrical Engineering and Computer Science
April 13, 2006

Certified by 4/13/06
Qing Hu
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Analysis of the Electron Transport Properties in Quantum Cascade Lasers

by

Hans Callebaut

Submitted to the Department of Electrical Engineering and Computer Science
on April 13, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Recently, the operating frequency range of quantum-cascade lasers (QCLs) has been extended from the mid-infrared to the far-infrared beow the Reststrahlen band (THz frequencies). Especially for THz QCLs, a detailed understanding of the dynamics of the electron transport is essential in order to extend their operation to longer wavelengths and higher temperatures. Compared to mid-infrared structures, the small subband separations in THz QCLs lead to LO-phonon scattering rates that are highly temperature sensitive and increases the importance of scattering processes which favor small transition energies, such as electron-electron (e-e), electron-impurity (e-imp)and electron-interface roughness scattering. This thesis details the development and calculation results for different models for the electron transport in THz QCLs.

Using a semi-classical Monte Carlo simulation, including e-e and e-LO-phonon scattering as well as e-imp scattering, the current density, population density and gain in two THz QCLs were investigated. We find that the inclusion of e-imp scattering in the calculations is crucial when modeling the intersubband transport dynamics in these devices. However, the calculated gain and current density exceed the measured values due to the absence of wavefunction localization and dephasing scattering in this model.

To describe coherent electron transport, a density matrix (DM) approach in combination with a tight-binding model was incorporated into the Monte Carlo simulation. The scattering events were treated semi-classically but contributed to dephasing scattering. In addition, a phenomenological "dephasing rate" was introduced to take into account dephasing caused by interface roughness scattering. This model was used to investigate the influence of dephasing on transport through a barrier. Additionally, current densities, populations and electron temperatures were calculated for two quantum-cascade structures and compared to a semi-classical simulation. We find that the inclusion of coherent transport and dephasing in the calculations is essential when transport is dominated by transitions between weakly coupled states.

As an alternative to the density matrix approach, a simulation based on the non-equilibrium Green's function formalism was implemented, based on the model described by Wacker [1]. We developed simulations which include momentum-dependent and momentum-independent scattering matrix elements. Scattering, including e-LO-phonon and e-imp scattering, is handled quantum-mechanically and takes into account the coherent interaction between states, as opposed to the semi-classical approach used in the DM picture. A model was developed for e-e scattering, which produced thermalized subband electron distributions. In addition, several quantum-cascade devices were simulated and the calculation results were compared to the DM Monte Carlo results and measurements. We find that, while the NEGF simulations with momentum-dependent scattering matrix elements and e-e scattering predict the most accurate results, the simple NEGF simulation originally implemented by Wacker and the DM simulation can also yield acceptable results at a much reduced computational expense.

Thesis Supervisor: Qing Hu
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to thank Professor Qing Hu for the opportunity to conduct this research and for providing the financial support that made it possible. I am very grateful for his confidence in my ability to to handle this project, and his useful suggestions when this work was taking shape. I would also like to thank him for the patience he has shown when things were not proceeding so smoothly. Further, I would like to thank Ben Williams and Sushil Kumar for the many helpful discussions we have had over the years. I would also like to thank Professor Goodnick of Purdue for providing us with his Monte Carlo code, which formed the framework we built our own MC simulations around. I am very grateful to Prof. Andreas Wacker of the University of Lund for letting me pick his brain about the NEGF simulations, and for his feedback and generous offer of his code. All the members in this group, including Ben Williams, Sushil Kumar, Alan Lee, Konstantinos Konistis and Qi Qin have supported and advised me in various ways, and I want to thank them. I would also like to mention Juan Montoya, for many extracurricular activities.

I would also like to thank my parents and family, for their continuing love and support. My presence here would not have been possible if it weren't for their taking care of so many things. I also want to thank the rest of my family and friends, for their support during difficult times. Last but not least, very special thanks to Katrien, who was taken away from us too soon. I couldn't have done this without you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The far-infrared frequency range is roughly defined as 30-300 μm or 4-40 meV. Often this range is also referred to by the term terahertz radiation, since 4-40 meV corresponds to 1-10 THz (Fig. 1-1). Far-infrared (FIR) or terahertz (THz) electromagnetic radiation is important in many applications such as radio astronomy, environmental monitoring, plasmon diagnostics, laboratory spectroscopy, telecommunications etc. and in the characterization of nanoscale condensed matter materials. In recent years, the generation, propagation and detection of FIR or THz electromagnetic radiation using two-dimensional semiconductor systems or other semiconductor nanostructures has become one of the most rapidly expanding fields in the photonics, optoelectronics and condensed matter physics communities.

Diode lasers are ideal sources because they are cheap, compact and very efficient. However, the semiconductor band gap places a limitation on emission frequency. The longest-wavelength diode lasers ($\sim 30\mu$m) are based on narrow gap lead-salt semiconductors [3]. While these lead-salt lasers have been quite successful for high resolution spectroscopy, they are still limited to cryogenic operation and provide relatively low power. On the other end of the spectrum, semiconductor transistors can be used to make 100 GHz oscillators [4]. Molecular gas lasers are currently the only practical laser sources for the far infrared, but they have limited lasing frequencies. They are

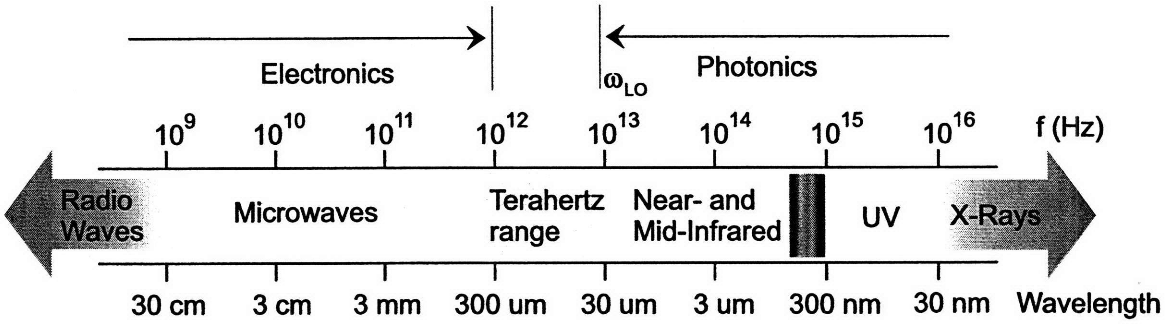

Figure 1-1: *The terahertz spectrum.*

also somewhat unwieldy as they require high voltage supplies and are usually rather bulky.

Intersubband lasers have several advantages over conventional semiconductor lasers. Most useful is the fact that the emission frequency is chosen by the design of the widths of the quantum wells, and can hence be tailored to the application. This is especially useful for infrared applications where small bandgap materials become difficult to find and work with. Also, since the envelope functions extend over a well (tens of Angstroms), the dipole moment for the intersubband transition is typically several orders of magnitude larger than that of an atomic transition. These features promise more efficient lasers.

1.2 Intersubband Lasers

In 1970, Esaki and Tsu [5] proposed using heterostructures for applications in opto-electronics. The use of intersubband transitions to create a laser was first suggested by Kazarinov and Suris [6] in 1971. Since then, electrically pumped quantum cascade lasers (QCLs) have been developed for wavelengths up to 161 μm [7]. Quantum wells are made by growing layers of different band gap semiconductors, such as GaAs and AlGaAs, on top of each other, creating a stack like structure. Since the bandgap of GaAs is smaller than that of $Al_xGa_{1-x}As$, the ensuing band gap profile gives rise to potential wells. The potential well height is determined by the Al alloy concentration of the barrier material.

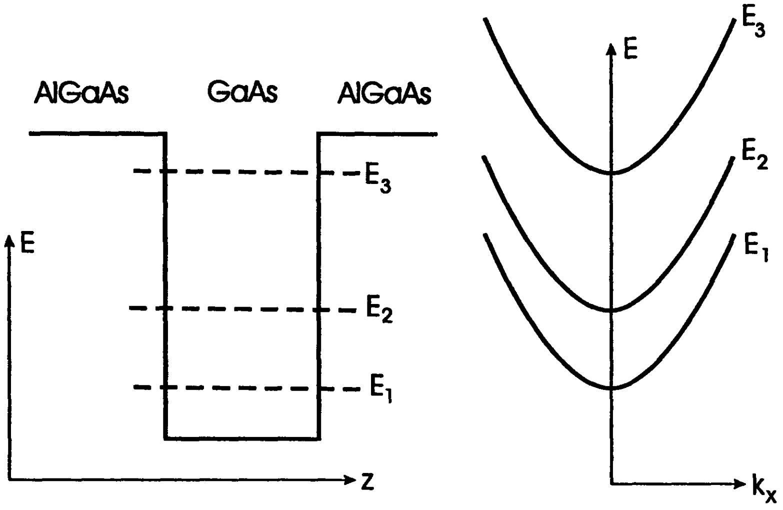

Figure 1-2: *Subbands in a quantum well. The potential well caused by the AlGaAs / GaAs quantum well gives rise to bound states localized in the well. In k-space, the subbands are parabolic as the electrons are not confined in the plane of the well.*

These quantum wells perturb the crystal periodicity in the growth direction. New electron energy states are located in these quantum wells, confined in the growth direction but still free in the plane of the well. As shown in figure (1-2), the conduction band is quantized into subbands.

The quantum well is similar to an impurity atom in that localized states are created. In momentum space, the subbands are parabolic as the electrons are not confined in the plane of the well. The exact energies of the subband minima are dependent on the well width and the depth of the potential well. The energies can be approximated by the formula for infinitely deep wells:

$$E_n = \frac{\hbar^2}{2m^*}(\frac{n\pi}{L})^2, \tag{1.1}$$

where m^* is the electron effective mass in GaAs, L is the well width and $\hbar$ is the reduced Planck's constant. The energy levels for a well with a finite barrier are lower than in equation (1.1). By choosing the well width and the barrier heights we can tailor the quantum levels so that transitions between E_m and E_n will emit photons in the far infrared. Very interesting from an engineering point of view is the ability to tune the energy levels and the dipole moments by applying a voltage bias. The

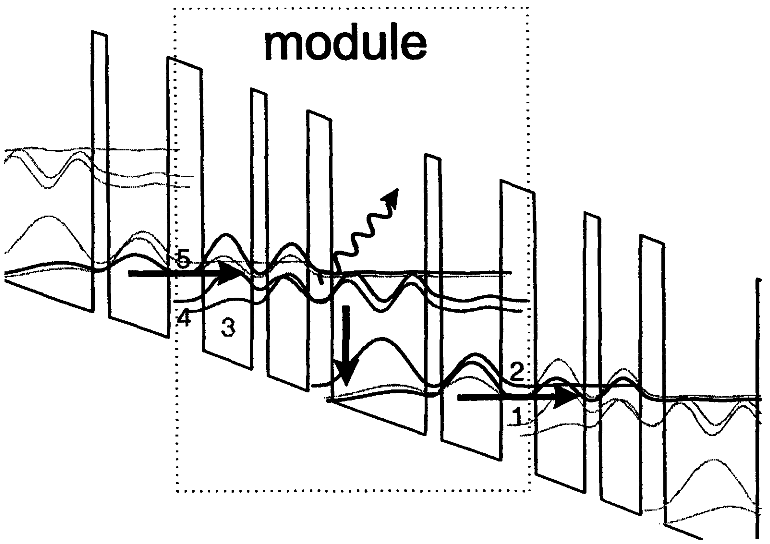

Figure 1-3: *Active region for the 3.4-THz QCL. Indicated are the conduction band profile and the magnitude squared of the wavefunctions. The arrows indicate the electron flow.*

Stark shift induced by the electric field shifts the energy levels and alters the potential profile. This is a very powerful tool when designing quantum wells.

Kazarinov and Suris [6] were the first to propose the use of intersubband transitions to design a laser. The first QCL was demonstrated by Federico Capasso in 1994 at Bell Labs for a wavelength of 4.2 μm, [8] and in 2001 Alessandro Tredicucci developed the first far-infrared QCL, operating at 68 μm. [9] QCLs have since achieved significant performance improvements and are poised to become the dominant laser sources in the mid- and far-infrared spectral ranges. Lasing has been obtained at wavelengths ranging from 3.4-24 μm [10, 8] in the mid-infrared and 60-161 μm in the far-infrared.

The general principles of how a QCL works are illustrated in figure 1-3 for the example of a 3.4-THz laser. [11] By applying a voltage bias, a potential staircase is created in which each identical step consists of multiple quantum wells, termed a "module". The operation of this device is based on a radiative transition between levels 5 and 4. Note that at terahertz frequencies, the non-radiative scattering lifetime τ_5 is typically much shorter (picoseconds) than the radiative lifetime (microseconds), so radiative transport plays no role below the lasing threshold. Depopulation of n=4 is achieved by its strong interaction with another level n=3, which is subject to fast resonant longitudinal optical phonon scattering into the collector/injector states n=2 and n=1. Electrons from these states are then injected into the excited radiative level of the next module, allowing the previously described process to repeat. A key feature of QCLs is the ability to cascade N modules together, so that a single injected electron can emit many photons, which allows for a differential quantum efficiency greater than unity.

Especially for THz QCLs, a detailed understanding of the dynamics of electron transport is essential in order to extend their operation to longer wavelengths and higher temperatures. Compared to mid-infrared structures, the subband separations in THz QCLs are much smaller, which greatly influences electron transport. In particular, since the radiative transition energy $\hbar\omega$ is smaller than the longitudinal optical (LO) phonon energy $\hbar\omega_{LO}$, the non-radiative relaxation rate is highly temperature dependent due to thermally activated LO-phonon scattering. The reduced subband separation also increases the importance of scattering processes which favor small transition energies, such as electron-electron (e-e), electron-impurity (e-imp), and interface roughness scattering. Therefore, in a proper analysis of electron transport in a THz QCL, all the aforementioned elastic and inelastic scattering mechanisms should be accounted for. Indeed, our calculations clearly indicate that electron-phonon (e-ph) and e-e scattering alone are insufficient to explain the measured current densities. [12]

Most QCL analyses and calculations [12, 13, 14, 15] have only considered e-ph and e-e scattering. Although the effects of impurity [16] and interface roughness scattering [17] on the spontaneous emission linewidth are well documented and generally accepted, their importance for electron transport in QCLs has been largely ignored. The study of e-imp scattering in the electron transport dynamics has been mostly restricted to the relaxation of excited carriers in quantum wells. [18, 19] In section 2.2.4, we show that the importance of e-imp scattering in electron transport in QCL rivals or even exceeds that of e-e scattering, and that it needs to be taken into account in a proper model of the electron transport dynamics.

Due to the often small intersubband energy separations in THz QCLs, the injection barrier in these devices is often quite thick to limit the interaction between the injector and the subbands in the next module. This design ensures that electrons will only be injected into the next module when the injector state lines up the excited radiative state in that module, and that parasitic currents are minimized. Therefore, resonant tunneling is a critical transport mechanism in THz QCLs, and as such it is the subject of active theoretical and experimental research. [20, 21, 22] However, while a qualitative understanding is straightforward, it is not always clear how to quantify the exact effect of coherent and incoherent transport. The effects of resonant tunneling and dephasing are most important when describing the transport between two weakly coupled energy states, e.g. tunneling through a thick barrier such as an injector barrier. In the calculation and analysis of QCLs, the localization of wavefunctions due to dephasing scattering is often disregarded, which can lead to unphysical results and limit the utility of the simulation. [23] Therefore, it is necessary to include a model for sequential tunneling to analyze the electron transport in QCLs over a broad bias range.

We investigated two different approaches to implement coherent transport in our simulations, namely the density matrix formalism and the non-equilibrium Green's function (NEGF) formalism. The density matrix (DM) formalism provides an easily accessible description of coherent electron transport, and is widely used to model optical and electronic transitions. [24] On the other hand, simulations using NEGF analysis [1] have shown promising results recently, but the complexity and computational burden of this method far exceed that of the DM approach and provide a less intuitive physical picture.

To assess the utility of a DM or NEGF model of electron transport in THz QCLs, simulation tools were developed for both approaches. This thesis describes the theory and implementation of these simulations, and compares the calculation results for the different models with each other and experimental measurements.

Chapter 2

Scattering and transport

2.1 Optical Transitions

The interaction of electromagnetic waves (light) and matter (electrons in the semiconductor) is the core of an optical device. Quantum mechanically, the interaction between photons and electrons in the semiconductor can be described by the Hamiltonian

$$H = \frac{1}{2m^*}(\mathbf{p} - q\mathbf{A})^2 + V(\mathbf{r}), \tag{2.1}$$

where $\mathbf{A}$ is the magnetic vector potential, m^* is the effective electron mass, and q is the carrier charge ($q = -e$ for electrons). Neglecting the term quadratic in $\mathbf{A}$ (a good approximation for most practical optical field intensities), and applying the Coulomb gauge $\nabla \cdot \mathbf{A} = 0$, we can distinguish the perturbation Hamiltonian H' due to the electron-photon interaction:

$$H' \approx \frac{e}{m^*}\mathbf{A} \cdot \mathbf{p}. \tag{2.2}$$

However, it is possible to take the quadratic term into account without making the perturbation Hamiltonian more cumbersome. We can do this by by explicitly writing the wavefunction ψ as the product of a phase factor and the remaining wavefunction ψ':

$$\psi \longrightarrow \psi' \, e^{ie\mathbf{A}\cdot\mathbf{r}/\hbar}. \tag{2.3}$$

Substituting this into the Schrödinger equation with Hamiltonian (2.1), we find:

$$\mathbf{p}\psi = e^{ie\mathbf{A}\cdot\mathbf{r}/\hbar}\,(\mathbf{p}+e\mathbf{A})\psi', \tag{2.4}$$

$$\frac{p^2}{2m^*}\psi' = i\hbar\left[\frac{\partial}{\partial t} - e\frac{\partial\mathbf{A}}{\partial t}\cdot\mathbf{r}\right]\psi', \tag{2.5}$$

and, as $\mathbf{E} = -\partial\mathbf{A}/\partial t$, the Schrödinger equation for the interaction of a photon with an electron can be written as a function of the amplitude of the incident optical field $\mathbf{E}$:

$$\left[\frac{p^2}{2m^*} - e\mathbf{E}\cdot\mathbf{r}\right]\psi' = i\hbar\frac{\partial\psi'}{\partial t}, \tag{2.6}$$

The perturbation Hamiltonian is:

$$H' = -e\mathbf{E}\cdot\mathbf{r}. \tag{2.7}$$

The physical interpretation of this interaction is more intuitively obvious than in the description with a vector potential. The radiative field acts as a force on the electron charge cloud, thus accelerating it and generating radiation (emitting a photon) or exciting the electron (absorption of a photon).

In order to describe the particle-particle interaction between photon and electron, we have to quantize the electric field E. This can be done similarly to the case of a harmonic oscillator. A photon then corresponds to one quantum of excitation in an oscillator:

$$E(\mathbf{r},t) = -i\sqrt{\frac{\hbar\omega}{2V\epsilon}}\left[\mathbf{a}^\dagger\, e^{-i\kappa\cdot\mathbf{r}+i\omega t} - \mathbf{a}\, e^{i\kappa\cdot\mathbf{r}-i\omega t}\right]. \tag{2.8}$$

The operators $\mathbf{a}$ and $\mathbf{a}^\dagger$ are photon annihilation and creation operators, respectively. They correspond to the absorption and emission of photons by mediation of an oscillating electric (electromagnetic) field with angular frequency ω. This is true even if that field is a vacuum field, as is the case for spontaneous emission.

The optical transition rate between an initial state $(E_i, \mathbf{k_i})$ and final state $(E_f, \mathbf{k_f})$ can be described using Fermi's Golden Rule:

$$W_{if} = \frac{2\pi}{\hbar}\left|\langle\psi_f|H'(\mathbf{r})|\psi_i\rangle\right|^2 \delta(E_f - E_i - \hbar\omega). \tag{2.9}$$

Here the delta function assumes a zero linewidth. In order to introduce a finite linewidth, the delta function can be replaced with the proper line-shape, usually a Lorentzian with linewidth Γ. The Lorentzian is a good model for line-shape broadening due to a finite lifetime or dephasing scattering.

$$\delta(E_f - E_i - \hbar\omega) \rightarrow \frac{\Gamma/(2\pi)}{(E_f - E_i - \hbar\omega)^2 + (\Gamma/2)^2}. \tag{2.10}$$

More generally, in most cases a number of final states is available, with density of states $\rho(E_{if})$. As each state is equally probable as a final state for the transition, we obtain (zero linewidth)

$$W_{if} = \frac{2\pi}{\hbar} \left|\langle\psi_f|H'(\mathbf{r})|\psi_i\rangle\right|^2 \rho(E_{if})\delta(E_f - E_i - \hbar\omega). \tag{2.11}$$

If we neglect non-parabolicity, the subbands in one particular band track each other. The energy separation between two states with identical in-plane wave vector remains constant for any two given subbands. Assuming only vertical transitions (dipole selection rule), this means that ρ will be given by the subband density of states for intersubband transitions.

We can write the initial (photon density n_{ph}) and final states ($n_{ph} + 1$) in an intersubband transition as:

$$\psi_i = \frac{e^{i\mathbf{k_{t,i}}\cdot\mathbf{r_t}}}{\sqrt{A}} F_i(z) u_{0,i}, \tag{2.12}$$

$$\psi_f = \frac{e^{i\mathbf{k_{t,f}}\cdot\mathbf{r_t}}}{\sqrt{A}} F_f(z) u_{0,f}, \tag{2.13}$$

where A is the in-plane area of the quantum well and $F(z)$ is an envelope function. As both the Bloch functions u belong to the same band, and as they are almost independent of $\mathbf{k}$, we can assume $u_{0,i} \approx u_{0,f}$. The matrix element H_{if} can then be written:

$$H_{if} \approx -i\sqrt{\frac{\hbar\omega(n_{ph}+1)}{2V\epsilon}}\ \hat{\mathbf{e}} \cdot \left\langle \frac{e^{i\mathbf{k_{t,f}}\cdot\mathbf{r_t}}}{\sqrt{A}} F_f(z) |e\mathbf{r}| \frac{e^{i\mathbf{k_{t,i}}\cdot\mathbf{r_t}}}{\sqrt{A}} F_i(z) \right\rangle \langle u_{0,f}|u_{0,i}\rangle, \tag{2.14}$$

$$H_{if} \approx -i\sqrt{\frac{\hbar\omega(n_{ph}+1)}{2V\epsilon}}\ \hat{\mathbf{e}} \cdot \langle F_f(z)|e\mathbf{z}|F_i(z)\rangle\ \delta_{\mathbf{k_{t,f}},\mathbf{k_{t,i}}}. \tag{2.15}$$

In the above equation we made use of the fact that the $F(z)$ can be considered a constant on the scale of a lattice spacing. This is a very good assumption for low level states and quantum wells wider than a few monolayers. This is generally true for the wave functions we are interested in.

The delta function in (2.15) corresponds to a conservation of in-plane momentum. The momentum carried by the photon, $\mathbf{k}_{ph}$, is of the order of $2\pi/\lambda$ ($\lambda \approx 100\mu$m), which is negligible compared to the electron wave vectors $\mathbf{k}_i$, $\mathbf{k}_f$ of the order of $2\pi/a$, a being the lattice constant (order of magnitude 5 Å). Therefore we can write $\mathbf{k}_{t,f} \approx \mathbf{k}_{t,i}$.

The matrix element $Z_{if} = \langle F_f|\mathbf{z}|F_i\rangle$ is called the dipole matrix element between the initial and final states. The dipole matrix element can be used as a gauge for the strength of the optical intersubband transition. Due to the dimensions of the quantum wells and their (bound) energy levels, Z_{if} in intersubband transitions ($\sim$ 30 Å) can be a lot larger than in an atomic system ($\sim$ 2 Å). During the design of a quantum well structure, we will try to maximize the dipole moment associated with the targeted intersubband transition.

Also apparent from (2.15) is a dipole selection rule for intersubband transitions. Only an electromagnetic wave with its electric field polarized along the z-axis (the quantum well growth axis) will be generated or absorbed in an intersubband transition.

Using Fermi's Golden Rule, the intersubband transition rate for stimulated emission (into one specific optical mode, i.e. the same one as the incident wave) can be written as:

$$W_{if} = \frac{\pi e^2 \omega Z_{if}^2 n_{ph}}{V\epsilon}\delta(E_f - E_i - \hbar\omega). \tag{2.16}$$

The transition rate is directly proportional to the intensity of the incident field ($\sim n_{ph}$). Equation 2.16 also shows that the transition rate decreases with increasing wavelength. The expression for (stimulated) absorption is identical to the one for stimulated emission.

For spontaneous intersubband emission, we have to sum over all available final photon states. Taking into account a 3D optical mode density of $(8\pi V n^3 E^2)/(h^3 c^3)$, the transition rate is:

$$W_{if,sp}^{3D} = \frac{e^2 n \omega^3 Z_{if}^2}{3\pi\epsilon_0 \hbar c^3}. \tag{2.17}$$

However, in far-infrared optical quantum electronic devices, the transition usually takes place inside a two-dimensional optical cavity with thickness t_c which is at the same scale or smaller than the wavelength (50-100 μm). This cavity can consist of a metal or plasma waveguide, confining the electromagnetic wave in the z-direction and limiting the optical mode density to $A/(2\pi)^2$. This yields a 2D intersubband transition rate of

$$W_{if,sp}^{2D} = \frac{e^2 n \omega^2 Z_{if}^2}{2 t_c \epsilon_0 \hbar c^2}, \tag{2.18}$$

scaling inversely proportional to the cavity thickness. Compared to the 3D expression, this dependence replaces a $1/\lambda$ dependence. This is shown more clearly if we look at the ratio of W^{3D} to W^{2D}

$$\frac{W_{if,sp}^{3D}}{W_{if,sp}^{2D}} = \frac{4t_c}{3\lambda}. \tag{2.19}$$

The microcavity effect will increase W^{2D} over the 3D case if the thickness of the cavity is smaller than the wavelength. Note that a microcavity only has an effect on the spontaneous emission rate. Stimulated emission, and hence gain, are not affected as all photons are coupled into one single mode. How many modes are available, is not important.

Optical gain is defined as the relative increase of a wave intensity per length unit as the wave propagates through the medium : $dI/dx = g(\omega)I$. To find the expression for optical gain, we subtract total absorption $\hbar\omega N_f W_{ab}$ from total stimulated emission $\hbar\omega N_i W_{st}$. With beam intensity $I = \frac{n_{ph}}{V}\hbar\omega\frac{c}{n}$ we find from (2.16):

$$W_{st} = \frac{e^2 n Z_{if}^2}{2\hbar^2 \epsilon c} I \delta(E_f - E_i - \hbar\omega), \tag{2.20}$$

and

$$g(\omega) = \frac{\Delta N e^2 \omega Z_{if}^2}{2\hbar n \epsilon_0 c}\delta(E_f - E_i - \hbar\omega). \tag{2.21}$$

Here $\Delta N = N_i - N_f$ is the population inversion between initial and final subbands. If the transition has a finite linewidth Δf, the delta function in (2.16) is replaced with a Lorentzian line-shape and the maximum gain is:

$$g_0 = \frac{\Delta N e^2 \omega Z_{if}^2}{\pi \hbar^2 n \epsilon_0 c \Delta f}. \tag{2.22}$$

2.2 Non-radiative transitions

2.2.1 Phonon scattering

The atoms in a semiconductor lattice are linked together with chemical bonds. These bonds can be strictly covalent or contain a degree of ionicity, as is the case between Ga(-) and As(+) in GaAs. Still, the atoms are constantly in motion, vibrating around their equilibrium lattice position, each atom a tiny harmonic oscillator. As the atomic vibrations are closely coupled through their common bonds, the atomic vibrations can be seen as part of larger lattice vibrations, which exist in several modes (see figure (2-1).

Similarly as with an electromagnetic field, each vibration mode can be quantized. A quantum of excitation in one mode is called a phonon, and each phonon can be characterized by a wave vector $\mathbf{q}$ and angular frequency ω. Like an electron or photon, an unconfined phonon can then be described by a (non-normalized) plane wave function $e^{i\mathbf{q}\cdot\mathbf{r}}$.

Similarly as for electrons, the lattice periodicity gives rise to a Brillouin-zone type $E - \mathbf{q}$ phonon dispersion relation. The lower branches represent the acoustic phonon modes, characterized by the neighboring atoms being in phase. In the longitudinal mode, the atomic displacements are in the same direction as the direction of energy transfer, while in the the transverse mode the atomic displacements are perpendicular to this direction. In optical phonon modes, the displacements of neighboring atoms are in opposite phase.

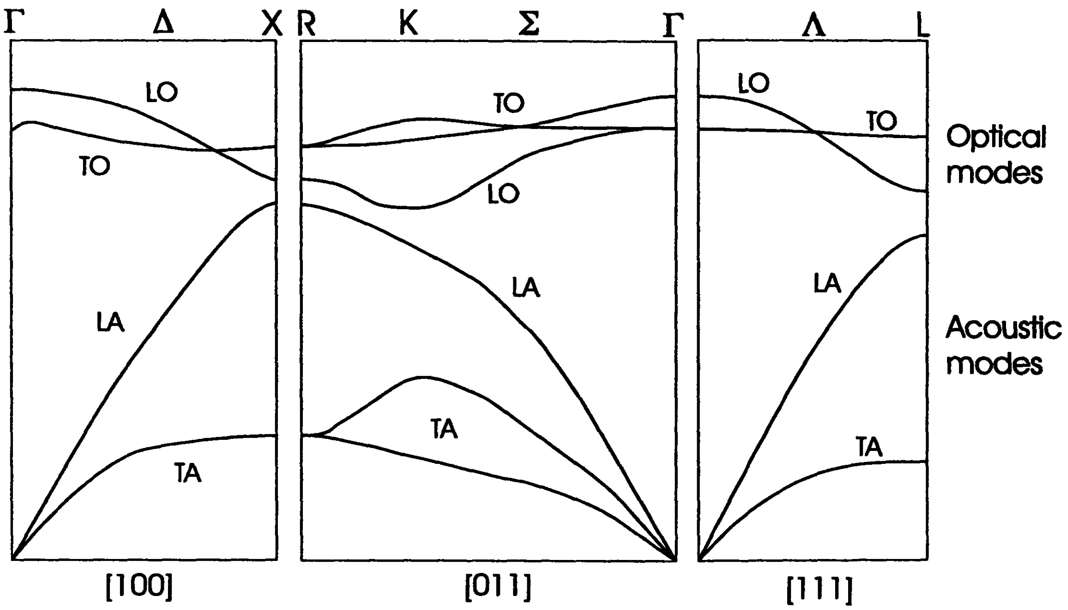

Figure 2-1: *Room temperature dispersion curves for acoustic and optical branch phonons in GaAs, obtained by inelastic neutron scattering. Adapted from Blakemore [2].*

As shown in the figure, the energy of the optical phonons is almost independent of $\mathbf{q}$, and in calculations longitudinal optical (LO) phonons are usually assumed to have one energy ($\hbar\omega_{LO}$ = 36 meV in GaAs). Because of the large dipole moment induced by neighboring ions, LO phonons couple strongly with electrons in polar semiconductors, provided enough energy is available.

As the phonons themselves represent the motion of atoms which are centers of charge, they also represent time-dependent perturbations of the crystal potential and can therefore scatter charge carriers. The electron-phonon interaction, i.e. creation and absorption of phonons, can be quantized through creation and absorption operators. The perturbation Hamiltonian H' is very similar to (2.8):

$$H' = \sum_{\mathbf{q}} \alpha(\mathbf{q}) \left[e^{-i\mathbf{q}\cdot\mathbf{r}}\, \mathbf{b_q}^{\dagger} + e^{i\mathbf{q}\cdot\mathbf{r}}\, \mathbf{b_q} \right]. \tag{2.23}$$

To find the total scattering rate, we sum over all $\mathbf{q}$ in the above equation. $\alpha(\mathbf{q})$ is the electron-phonon coupling strength.

LA phonons are most important for low energies or low temperatures. In these cases they correspond to long wavelength deformations of the crystal lattice. The interaction strength can be expressed as [25]:

$$|\alpha(\mathbf{q})|^2 = \frac{\hbar\omega D^2}{2\rho V c_s^2}. \tag{2.24}$$

Here D refers to the deformation potential, V is the crystal volume, ρ the material density and c_s the (longitudinal) speed of sound in the material.

For polar semiconductors such as GaAs, interactions with LO phonons are most important. The LO phonon interaction strength is [26]:

$$|\alpha(\mathbf{q})|^2 = \frac{\hbar\omega_{LO} e^2(\epsilon_s - \epsilon_\infty)}{2\epsilon_s\epsilon_\infty\epsilon_0 V q^2}, \tag{2.25}$$

where ϵ_s and ϵ_∞ denote the relative permittivity of GaAs at frequencies lower and higher than optical frequencies, respectively. Due to the $1/q^2$ dependence, interactions with LO phonons at the zone center are favored over transitions involving a large momentum transfer. Also, the LO phonon threshold energy (36 meV for GaAs) causes a sharp temperature dependence. This effect is especially important for far-infrared transitions, where the subband energy spacing is less than $\hbar\omega_{LO}$. Here, hot carriers can open up parasitic LO phonon channels, drastically altering average scattering times. This can be detrimental to the working of the device.

We can adapt the bulk phonon expression to the 2D case by splitting the real space dependence of the Hamiltonian in components along and perpendicular to the growth axis.

$$H' = e\sum_{\mathbf{q}} \left[\frac{\hbar\omega_{LO} e^2(\epsilon_s - \epsilon_\infty)}{2\epsilon_s\epsilon_\infty\epsilon_0 V q^2}\right]^{1/2} \frac{e^{-i\mathbf{q_t}\cdot\mathbf{r_t}}}{\sqrt{A}} \frac{e^{-iq_z z}}{\sqrt{L}}. \tag{2.26}$$

To assess the transition rate between an initial state $\mathbf{k}_i$ on subband i and a final state $\mathbf{k}_f$ on subband f, we use Fermi's Golden Rule:

$$W_{\mathbf{k}_i\rightarrow\mathbf{k}_f} = \frac{2\pi}{\hbar} H'_{if}\, \delta(E_i(\mathbf{k}_i) - E_f(\mathbf{k}_f) - \hbar\omega(\mathbf{q})). \tag{2.27}$$

We can rewrite the matrix element H'_{if}:

$$H'_{if} = \alpha(q) A_{if}(q_z)\, \delta_{\mathbf{k}_i - \mathbf{k}_f, \mathbf{q}_t}, \tag{2.28}$$

where $A_{if} = \int_{-\infty}^{+\infty} \psi_f{}^\dagger(z)\, \psi_i(z)\, e^{iq_z z} dz.$

The form factor A_{if} contains the dependence on the electron wave functions. Summing over all possible $\mathbf{k}_f$, we find the total rate for a LO phonon mediated transition from the initial state $\mathbf{k}_i$ in subband i to a state in subband f :

$$W_{\mathbf{k}_i} = \frac{2\pi}{\hbar} \sum_{\mathbf{k}_f} \sum_{\mathbf{q}_t} \sum_{q_z} |\alpha(q)|^2 \, |A_{if}(\mathbf{q})|^2 \, \delta(E_i - E_f - \hbar\omega(\mathbf{q})). \tag{2.29}$$

We assume parabolic subbands and $\hbar\omega(\mathbf{q}) \approx \hbar\omega_{LO}$. The electrons in subband i are thermalized with electron temperature $T_{e,i}$, and their energy distribution can be described using a Fermi-Dirac distribution around a chemical potential (Fermi energy) E_F:

$$f(E) = \frac{1}{e^{\frac{E-E_F}{kT_{e,i}}} + 1}. \tag{2.30}$$

As phonons are bosons, their energy distribution is the Bose-Einstein function:

$$N_{LO}(E) = \frac{1}{e^{\frac{E-\mu_{LO}}{kT_{ph}}} - 1}. \tag{2.31}$$

Using the momentum conservation $\mathbf{q}_t = \mathbf{k}_i - \mathbf{k}_f$ from (2.28) the LO phonon scattering rate can be written:

$$\begin{aligned} W_{k_i} &= \frac{e^2(\epsilon_s - \epsilon_\infty)\omega_{LO}\, m^*}{4\pi^2\epsilon_0\epsilon_\infty\epsilon_s\hbar^2} \int d\phi \int_0^{+\infty} dq_z \, (1 - f_f(E_f)) \frac{|A_{if}(q_z)|^2}{q^2} \\ &\quad \times \, (N_{LO}(\mathbf{q}) + 1) \, H(E_i^k + \Delta E_{if} - \hbar\omega_{LO}). \end{aligned} \tag{2.32}$$

E_i^k is the kinetic energy of the electron in the initial subband, $E_i^k = \hbar^2 k_{i,t}^2/2m^*$. The energy to create a phonon can be obtained from both the electron kinetic energy and potential energy, i.e. the subband separation. This is reflected in the step function $H(E_i^k + \Delta E_{if} - \hbar\omega_{LO})$. It is possible to transform the expression for the scattering rate by Fourier transforming the form factor:

$$\begin{aligned} W_{k_i} &= \frac{e^2(\epsilon_s - \epsilon_\infty)\omega_{LO}\, m^*}{4\pi\epsilon_0\epsilon_\infty\epsilon_s\hbar^2} \int d\theta \, (1 - f_f(E_f)) \frac{|F_{if}(q_t)|}{q_t} \\ &\quad \times \, (N_{LO}(\mathbf{q_t}) + 1) \, H(E_i^k + \Delta E_{if} - \hbar\omega_{LO}). \end{aligned} \tag{2.33}$$

with

$$F_{if} = \int dz \int dz' \, \psi_f{}^\dagger(z)\psi_i(z)\psi_f(z')\psi_i{}^\dagger(z'), \tag{2.34}$$

and θ is the angle between $\mathbf{q_t}$ and the x-axis.

Usually we can assume that the phonon temperature T_{ph} is close to the lattice temperature T_l. For experiments where the device is being tested in a cryogenic environment, mounted on a cold plate, this means that LO phonons are frozen out. The equilibrium LO phonon population is negligible and scattering is dominated by the phonon emission process.

However, resonant LO-phonon scattering can result in the generation of an abundance of phonons with a very similar momentum. Due to this non-equilibrium phonon population, the reabsorption of the hot LO-phonons can become significant. For a resonant transition (τ_{LO} ≈0.3 ps) from an upper subband with a population density n_i, most phonons are generated with a momentum $\mathbf{q} \approx \mathbf{0}$, and the generation rate of LO-phonons R_{LO} can be approximated by:

$$R_{\mathrm{LO}} = \frac{n_i}{\tau_{LO}}. \tag{2.35}$$

For an electron gas with a temperature of 60 K (average $E_i^k = kT_{e,i} \approx 5$ meV), an electron with kinetic energy E_i^k will be scattered by a LO-phonon with a momentum q_t ranging between 0 and $2k_i = 2\sqrt{2m^* E_i^k}/\hbar \approx 2 \times 10^8\ \mathrm{m}^{-1}$. However, LO-phonon scattering is much faster for small q_t due to the $1/q_t$ dependence in Eq. 2.33 so the contributions to the total scattering rate from phonons with large q_t can generally be neglected compared to small q_t. To simplify the problem, we assume the phonons are spread uniformly over momentum space with $q_t < q_t^{\mathrm{max}} = 1 \times 10^8\ \mathrm{m}^{-1}$. For a LO-phonon dissociation time $\tau_{\mathrm{LO}}^{\mathrm{diss}} \approx 5$ ps, the non-equilibrium phonon occupation number $N_{\mathrm{LO}}^{\mathrm{non-eq}}$ is given by:

$$N_{\mathrm{LO}}^{\mathrm{non-eq}} = R_{\mathrm{LO}} \tau_{\mathrm{LO}}^{\mathrm{diss}} \frac{(2\pi)^2}{\pi (q_t^{\mathrm{max}})^2} \tag{2.36}$$

Absorption of LO-phonons will become important if it rivals LO-phonon emission, i.e. if $N_{\mathrm{LO}}^{\mathrm{non-eq}} \approx 1$ which happens for a population density of

$$n_i = \frac{(2\pi)^2}{\pi q_{\mathrm{max}}^2} \frac{\tau_{\mathrm{LO}}}{\tau_{\mathrm{LO}}^{\mathrm{diss}}} = 4.7 \times 10^9 \mathrm{cm}^{-2}. \tag{2.37}$$

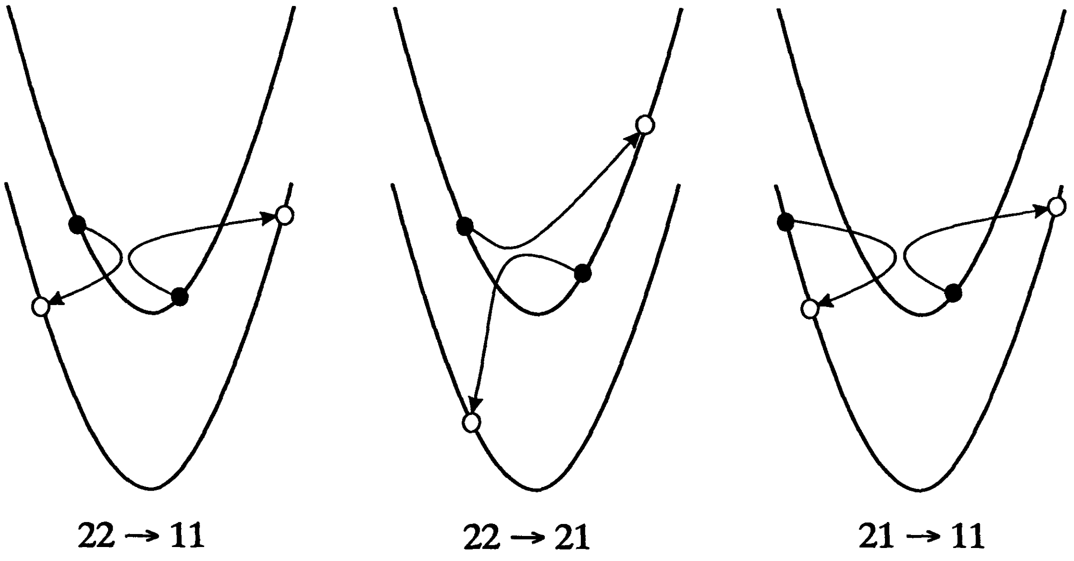

Figure 2-2: *Various intersubband carrier-carrier scattering mechanisms for a two-subband system*

2.2.2 Electron-electron scattering

With increasing population density, electrons are more and more likely to interact and scatter. Especially in cases where LO-phonon scattering is not possible or very limited, e-e scattering is the main scattering mechanism. In this section we will be using the Hartree approximation, in which we neglect the "exchange energy" caused by the anti-symmetry in the real space wave function of a two-electron state if the two electrons have the same spin. Inclusion of this "exchange energy" adds considerable complexity to the problem (Hartree-Fock), effectively making the problem intractable.

The perturbation Hamiltonian is an unscreened Coulombic potential:

$$H' = \frac{e^2}{4\pi\epsilon r}, \tag{2.38}$$

with r the distance between the electrons and $\epsilon = \epsilon_s\epsilon_0$ the dielectric permittivity of the semiconductor. At its simplest, we can represent electron-electron interaction as a two-body process involving two isolated carriers. As there are two initial and two final states, there are a lot more scattering possibilities than in the case of LO-phonon scattering, which involved one initial and one final state. In figure (2-2), various scattering mechanisms for intersubband scattering are illustrated. The transition

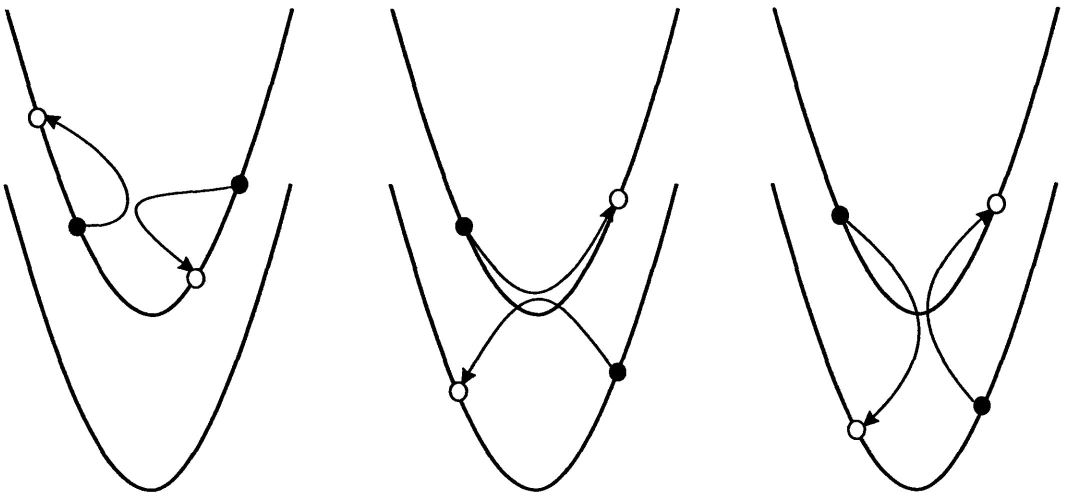
22 → 22 21 → 21 21 → 12

Figure 2-3: *Various intrasubband carrier-carrier scattering mechanisms for a two-subband system*

from subband 2 to subband 1 can be split into three contributions, 22-11, 22-21 and 21-11. The 22-21 and 21-11 transitions are Auger-type transitions, with one electron relaxing down to a lower subband while giving its excess energy to another electron which scatters higher into its original subband. Also, there are scattering events which don't affect the number of electrons in a subband, as illustrated in figure (2-3). 22-22 is a "pure" intrasubband scattering event. Even though these intrasubband e-e scatterings don't change the subband populations, they are very important for thermal equilibrium in and between subbands.

The initial and final states are composed of two electron wavefunctions, and are of the form $|\psi_{12}\rangle = |\psi_1\rangle|\psi_2\rangle$. The collision probability between electrons with equal spin polarity is lower due to an exchange term (related to Pauli's exclusion principle), and therefore only electrons with opposite spins are taken into account here [27]. Taking plane waves for the electron wavefunctions, the matrix element H'_{if} becomes:

$$H'_{if} = \left\langle \psi_f(z)\frac{e^{-i\mathbf{k}_f\cdot\mathbf{r}_t}}{\sqrt{A}}\psi_g(z')\frac{e^{-i\mathbf{k}_g\cdot\mathbf{r}'_t}}{\sqrt{A}} \middle| \frac{e^2}{4\pi\epsilon r} \middle| \psi_i(z)\frac{e^{-i\mathbf{k}_i\cdot\mathbf{r}_t}}{\sqrt{A}}\psi_j(z')\frac{e^{-i\mathbf{k}_j\cdot\mathbf{r}'_t}}{\sqrt{A}} \right\rangle, \tag{2.39}$$

where the initial electron states are labeled i and j, and the final states f and g. Note that, for simplicity, we are working with the unscreened Coulombic potential. The separation of the carriers is:

$$r = \sqrt{\left|\mathbf{r}_t - \mathbf{r}'_t\right|^2 + (z - z')^2}. \tag{2.40}$$

Therefore we obtain:

$$\begin{aligned} H'_{if} &= \frac{e^2}{4\pi\epsilon A^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int \int \psi_f{}^{\dagger}(z)\, \psi_g{}^{\dagger}(z')\, \psi_i(z)\, \psi_j(z') \\ &\quad \times \frac{e^{i(\mathbf{k}_i\cdot\mathbf{r}_t + \mathbf{k}_j\cdot\mathbf{r}'_t)}\, e^{-i(\mathbf{k}_f\cdot\mathbf{r}_t + \mathbf{k}_g\cdot\mathbf{r}'_t)}}{\sqrt{\left|\mathbf{r}_t - \mathbf{r}'_t\right|^2 + (z - z')^2}}\, d\mathbf{r}_t\, d\mathbf{r}'_t\, dz\, dz'. \end{aligned} \tag{2.41}$$

Expanding the Coulombic potential in a Fourier series, and substituting H'_{if} into Fermi's Golden Rule gives the scattering rate of a carrier in subband i. Integrating over all the states of the second carrier (given by $\mathbf{k}_j$) and introducing Fermi-Dirac distribution functions to account for state occupancy, we find

$$\begin{aligned} W &= \frac{e^4}{2\pi\hbar(4\pi\epsilon)^2} \int\int\int \frac{|A_{ijfg}(q_t)|^2}{q_t^2} f_j(\mathbf{k}_j)\,[1 - f_f(\mathbf{k}_f)]\,[1 - f_g(\mathbf{k}_g)] \\ &\quad \times \delta(\mathbf{k}_f + \mathbf{k}_g - \mathbf{k}_i - \mathbf{k}_j)\; \delta(E_f^t + E_g^t - E_i^t - E_j^t)\; d\mathbf{k}_g\, d\mathbf{k}_i\, d\mathbf{k}_j. \end{aligned} \tag{2.42}$$

where the energies E^t refer to the total energy of the corresponding carrier, i.e. subband minimum energy plus kinetic energy. A_{ijfg} is a form factor and a function of $q_t = |\mathbf{k}_i - \mathbf{k}_f|$:

$$A_{ijfg}(q_t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \psi_f{}^{\dagger}(z)\, \psi_g{}^{\dagger}(z')\, \psi_i(z)\, \psi_j(z') e^{-iq_t|z-z'|}\, dz'\, dz. \tag{2.43}$$

The delta functions express the conservation of momentum and energy in the scattering event. We can see that carrier-carrier scattering will be largest for small exchanged wavevectors. Assuming parabolic subbands with $E^t = E + \hbar^2 k^2/2m^*$, we find:

$$\begin{aligned} W &= \frac{m^* e^4}{\pi\hbar^3(4\pi\epsilon)^2} \int\int \frac{|A_{ijfg}(q_t)|^2}{q_t^2} P_{j,f,g}(\mathbf{k}_j, \mathbf{k}_f, \mathbf{k}_g) \\ &\quad \times \delta\Big(\mathbf{k}_f{}^2 + \mathbf{k}_g{}^2 - \mathbf{k}_i{}^2 - \mathbf{k}_j{}^2 + \frac{2m^*}{\hbar^2}(E_f + E_g - E_i - E_j)\Big)\, d\mathbf{k}_f\, d\mathbf{k}_j, \end{aligned} \tag{2.44}$$

with $P_{j,f,g}(\mathbf{k}_j, \mathbf{k}_f, \mathbf{k}_g)$ representing the probability functions.

It is useful to introduce two new variables, the relative wavevectors

$$\mathbf{k}_{ij} = \mathbf{k}_j - \mathbf{k}_i, \tag{2.45}$$

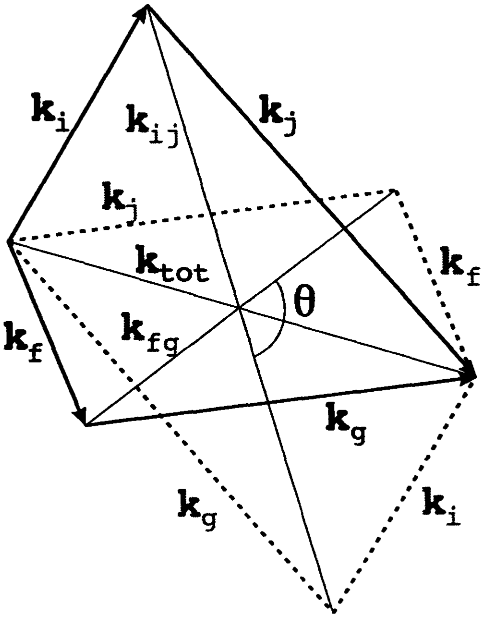

Figure 2-4: *Conservation of momentum in e-e scattering.*

$$\mathbf{k}_{fg} = \mathbf{k}_g - \mathbf{k}_f. \tag{2.46}$$

The energy conserving delta function then allows reduction of this integral to:

$$W = \frac{m^* e^4}{\pi \hbar^3 (16\pi\epsilon)^2} \int \int_0^{2\pi} \frac{|A_{ijfg}(q_t)|^2}{q_t^2} P_{j,f,g}(\mathbf{k}_j, \mathbf{k}_f, \mathbf{k}_g) d\theta \, d\mathbf{k}_j, \tag{2.47}$$

and θ is the angle between $\mathbf{k}_{ij}$ and $\mathbf{k}_{fg}$, as illustrated in figure (2-4). Although (2.47) looks simple, the actual computation is rather time-consuming and resource-intensive. Certain simplifications, like ignoring final-state Fermi blocking, are common. This is a fair approximation for low carrier densities or high electron temperatures.

At higher electron densities, the interaction can no longer be described as solely between two isolated carriers. The reaction of other carriers to the Coulombic potential will effectively "screen" the disturbing field, thereby reducing the perturbation. The probability of scattering will decrease as compared to the non-screened case.

One of the simplest models for screening considers only the carriers within the same subband as the initial carrier state. It replaces [28] the dielectric constant ϵ_s with one which is dependent upon the relative wave vector q_t:

$$\epsilon_s = 1 + \frac{2\pi e^2}{(4\pi\epsilon_s) q_t} \Pi_{ii}(q_t, T) A_{ijfg}(q_t), \tag{2.48}$$

with the polarization factor

$$\Pi_{ii}(q_t, T) = \int_0^{+\infty} \frac{\frac{m^*}{\pi\hbar^2}\left[1 - \mathrm{H}(q_t - 2k_F)\sqrt{1 - \left(\frac{2k_F}{q_t}\right)^2}\right]}{4kT\cosh^2(\frac{E_F - E}{2kT})}\, dE. \tag{2.49}$$

k_F is the Fermi wave vector for subband i.

Equation (2.47) gives the carrier-carrier scattering rate for a particular carrier energy i, averaged over another initial carrier distribution j. In order to find a scattering rate for the whole subband i, we have to average out over the Fermi-Dirac distribution of carriers in the initial state.

$$\frac{1}{\tau} = \frac{\int \frac{1}{\tau_i} f_i(E_i^k)\, dE_i^k}{\int f_i(E_i^k)\, dE_i^k}, \tag{2.50}$$

with E_i^k the kinetic energy associated with $\mathbf{k}_i$. The denominator is equal to $N_i\pi\hbar^2/m^*$, and assuming a parabolic subband, we obtain:

$$\frac{1}{\tau} = \frac{\int \frac{1}{\tau_i} f_i(k_i) k_i\, dk_i}{\pi N_i}. \tag{2.51}$$

In general, the "intrasubband" carrier scattering rate increases with temperature, as state blocking becomes less important. The rise in the number of easily accessible final states results in a higher scattering rate.

For intersubband scattering mechanisms this is less evident. Overall, intersubband e-e scattering can be considered to be nearly temperature insensitive. For the population densities commonly encountered in far-IR QCLs, the intersubband e-e scattering rate is approximately proportional to the upper subband population. State blocking is usually not an issue, again due to the low population densities.

Smet [26] showed that intersubband e-e scattering is nearly inversely proportional to the intersubband energy separation. However, the relation becomes more complex if the two levels considered are close to anti-crossing. The overlap between the wavefunctions of initial and final states will sensitively depend on their energy difference.

We can summarize the above as:

$$\frac{1}{\tau} \propto \frac{N}{\Delta E_{21}}. \tag{2.52}$$

This empirical relation can be used as a rough guideline in the design of quantum well structures where carrier-carrier scattering plays an important role.

2.2.3 Impurity Scattering and Interface Roughness Scattering

The background concentration of ionized impurities, imbedded in the semiconductor lattice, and its associated distributed charges cause coulomb interaction and scattering with the transport electrons in the device. In order to investigate and quantize this scattering, we assume that the impurity density is low enough so that each individual impurity can be treated separately from the others.

The scattering potential of an ionized impurity at location $z = z_{\mathrm{imp}}$ (z-direction is the growth direction) is given by the Coulomb potential, and we can write for the corresponding matrix element describing a transition between an initial state i to a final state f:

$$\langle f|H^{\mathrm{imp}}|i\rangle = \frac{e^2}{2\epsilon A q} A_{if}^{\mathrm{imp}}(q, z_{\mathrm{imp}}), \tag{2.53}$$

where A is a generalized area, $\mathbf{q} = \mathbf{k_f} - \mathbf{k_i}$ is the exchanged momentum. The form factor A_{if}^{imp} is given by:

$$A_{if}^{\mathrm{imp}}(q, z_{\mathrm{imp}}) = \int dz \phi_f(z) \phi_i(z) \mathrm{e}^{-q|z - z_{\mathrm{imp}}|} \tag{2.54}$$

Using Fermi's Golden Rule, we can then find the scattering rate for an electron from subband i and momentum $\mathbf{k}_i$ to a final subband f by summing over the all possible final momenta $\mathbf{k}_f$:

$$\begin{aligned} W_{i\to f}^{\mathrm{imp}} &= \frac{2\pi}{\hbar} \sum_{\mathbf{k}_f} |H_{if}|^2 \delta(E_f - E_i) \\ &= \frac{2\pi}{\hbar} \sum_{\mathbf{k}_f} \left(\frac{e^2}{2\epsilon A q} \right)^2 \delta(E_f - E_i) \end{aligned} \tag{2.55}$$

If we transform the sum into an integral and take into account spin conservation, we can write the scattering as:

$$W_{i\to f}^{\mathrm{imp}} = \frac{m^* e^4}{8\pi \hbar^3 \epsilon^2 A} \int_0^{2\pi} \left| \frac{A_{if}^{\mathrm{imp}}(q, z_{\mathrm{imp}})}{q} \right|^2 d\theta, \tag{2.56}$$

where m^* is the effective electron mass and θ is the angle between $\mathbf{k}_f$ and the x-axis (in-plane). Note that this description does not refer to the in-plane coordinates of the ionized impurity, which is due to the delocalized nature of the electron wavefunctions in those directions. Since the scattering rate only depends on z_{imp} and no other coordinates, it is very straightforward to generalize Eq. 2.56 to describe the effects of a sheet charge $N(z_{\text{imp}})dz$ at $z = z_{\text{imp}}$:

$$W^{\text{imp}}_{i\to f}(z_{\text{imp}})dz = \frac{m^* e^4 N(z_{\text{imp}})dz}{8\pi\hbar^3\epsilon^2 A}\int_0^{2\pi}\left|\frac{A^{\text{imp}}_{if}(q, z_{\text{imp}})}{q}\right|^2 d\theta \tag{2.57}$$

For a bulk doping described by a distribution $N(z)$ the total scattering rate is given by:

$$W^{\text{imp}}_{i\to f}(z_{\text{imp}}) = \int dz_{\text{imp}} N(z_{\text{imp}})\frac{m^* e^4}{8\pi\hbar^3\epsilon^2 A}\int_0^{2\pi}\left|\frac{A^{\text{imp}}_{if}(q, z_{\text{imp}})}{q}\right|^2 d\theta. \tag{2.58}$$

Monolayer fluctuations of the barrier and quantum well thicknesses results in a position dependent variation of the energy and wavefunction of the subbands. [29] We assume that the roughness height $\Delta(\mathbf{r})$ at the in-plane position $\mathbf{r}$ has a correlation function:

$$\langle\Delta(\mathbf{r})\Delta(\mathbf{r}')\rangle = \pi\frac{\Lambda^2\Delta^2}{A}\left(1 + \frac{q^2\Lambda^2}{2}\right)^{-3/2}, \tag{2.59}$$

where A is the interface area, Δ is the mean height of roughness and Λ is the correlation length. The scattering matrix element is given by:

$$\langle f\mathbf{k}' H^{\text{rough}} || i/momk\rangle = \int d^2r\, F_{if}\Delta(\mathbf{r})\, e^{\mathbf{i}\cdot\mathbf{r}} \tag{2.60}$$

with

$$F_{if} = V_0\phi_i(z_0)\phi_f(z_0), \tag{2.61}$$

where V_0 is the barrier height (approximately 125 meV for GaAs/$\text{Al}_{0.15}\text{Ga}_{0.85}$Asstructures) and $\phi_i(z_0)$ is the wavefunction of subband i at the interface. Because interface roughness is equivalent to local fluctuations in well width, F_{if} can also be expressed as:

$$F_{if} = \sqrt{(\partial E_i/\partial L)(\partial E_f/\partial L)}, \tag{2.62}$$

where L is the well width. From Fermi's Golden Rule we find the scattering rate from state $|i, \mathbf{k}\rangle$ to $|f, \mathbf{k}'\rangle$:

$$W^{\text{rough}}_{i\to f} = \frac{2\pi}{A\hbar}V_0^2\pi\Delta^2\Lambda^2\left(1 + \frac{q\Lambda^2}{2}\right)^{-3/2}\delta(E_f - E_i)\left\langle\phi_f(z_0)\phi_i(z_0)\right\rangle. \tag{2.63}$$

Note that this scattering rate is proportional with the square of the correlation length Λ^2, and that for a perfectly flat surface $\Lambda = \infty$ the scattering rate approaches infinity. However, physically this corresponds with a "glancing" scattering process in which no momentum or energy are exchanged, and the scattering event has no significance because no observable parameters are affected. However, this infinity causes some problems when this expression for the scattering rate is used in simulations. Usually, we only consider scattering processes with a certain minimum exchanged momentum (typically q ¿ 2×10^7 m^{-1}).

2.2.4 Relative importance of impurity and e-e scattering

Ionized impurities (and interface imperfections) are static scatterers, and therefore it is always possible to rediagonalize the Hamiltonian to get stationary wavefunctions that take into account the modified potential landscape. In such a picture, the imperfections would not cause any inter-eigenstate scattering per se. The in-plane translational symmetry would be destroyed, and transitions would take place between manifolds of states with in-plane position dependence, resulting in linewidth broadening. However, in simulations a perturbative approach (treating impurities and imperfections as random scatterers in an otherwise perfect lattice) is preferable. This approach allows us to describe the in-plane component of the wavefunction as a plane wave, which vastly reduces the calculation's complexity and preserves an intuitive picture of intersubband transitions.

Interface roughness scattering rates depend sensitively on the details and conditions of the growth, and will vary widely with samples. This makes interface roughness scattering impossible to quantify in a universal way without introducing phenomenological parameters. In contrast, the distribution of ionized impurities in a sample can be modeled accurately. Due to these considerations, we have chosen to focus on e-imp scattering. As will be shown, its importance in intersubband transport rivals or even exceeds that of e-e scattering.

Both e-imp and e-e scattering are Coulombic interactions, which allows for a simple assessment of their relative importance. Assuming charge neutrality, the number of electrons and ions are the same and hence the number of possible scattering centers is equal for both processes. However, due to the exchange interaction, e-e scattering is mostly caused by interactions between electrons of opposite spins. [27] Also, in a center-of-mass frame, e-e scattering can be described with a reduced mass $m_r^* = m^*/2$, whereas for e-imp scattering $m_r^* = m^*$. Furthermore, intersubband e-e scattering largely originates from interactions between electrons in the same subband. Impurity scattering, on the other hand, is limited by its dependence on the distance $|z - z_{imp}|$ between the scattering electron and the ionized impurity, and as such is most effective for transport between subbands whose wavefunctions $\chi(z)$ are close to the doping layer. This rough estimate of the relative importance of e-e and e-imp scattering [19] can be summarized as:

$$\frac{W_{if}^{\mathrm{e-e}}}{W_{if}^{imp}} \propto \frac{1}{4} \frac{N_i |A_{iiff}^{\mathrm{e-e}}(q)|^2}{N_{imp} |A_{if}^{\mathrm{imp}}(q)|^2}. \tag{2.64}$$

Here $W_{if}^{\mathrm{e-e}}$ and W_{if}^{imp} are, respectively, the e-e and e-imp scattering rates from $n = i$ into $n = f$, N_i is the population density in subband i, N_{imp} is the total doping density, and the form factors A (Eqs. 2.43 and 2.54) are functions of the exchanged momentum q. The factor-of-four reduction arises from the exchange interaction and reduced mass discussed above. Clearly, for intersubband transport, e-imp scattering usually dominates over e-e scattering and adds significantly to electron gas heating. However, e-imp scattering does not allow for energy exchange between electrons, and therefore does not contribute to intrasubband carrier thermalization as e-e scattering does.

2.3 Coherent transport

2.3.1 General description

In quantum cascade lasers (QCLs), resonant tunneling is a critical transport mechanism, and as such it is the subject of active theoretical and experimental research. [20, 21, 22] However, while a qualitative understanding is straightforward, it is not always clear how to quantify the exact effect of coherent and incoherent transport. The effects of resonant tunneling and dephasing are most important when describing the transport between two weakly coupled energy states, i.e. tunneling through a thick barrier such as an injector barrier. In the calculation and analysis of QCLs, the localization of wavefunctions due to dephasing scattering is often disregarded, which can lead to unphysical results and limit the usefulness of the simulation. [23] Therefore, it is necessary to include a model for sequential tunneling to analyze the electron transport in QCLs over a broad bias range. Although simulations using non-equilibrium Green (NEG) function analysis [1], as described in a following chapter, are showing promising results, the complexity and computational burden of this method limit its utility in obtaining an intuitive picture of electron transport. On the other hand, the density matrix formalism provides an easily accessible description of coherent electron transport, and is widely used to model optical and electronic transitions. [24] A more comprehensive description of this formalism is provided in the next section.

The importance of coherent transport in multiple quantum well structures can be appreciated by considering a simple superlattice as shown in Fig. 2-5. In this simple structure, only two energy levels in each well, 1 and 2, participate in the transport process as electrons move from the left to the right under an applied electric field. Fig. 2-5(a) illustrates the scheme that is described by the semi-classical model, in which the entire superlattice is treated as a single quantum mechanical system with a well-defined Hamiltonian. All the subband energy levels are eigenstates (which are stationary by definition) of this Hamiltonian. The transport process is the collective effect of intersubband scattering between the various subbands (eigenstates) involved, and can be calculated using the Fermi's Golden Rule approximation. This is essentially a rate-equation approach and there is no coherent oscillatory time evolution among the subband electron populations. In this picture, the electron wavefunctions always correspond to the stationary eigenstates, and scattering transports an electron from one eigenstate to another. Under resonant bias conditions, the ground state $1'$ in one well is aligned with the excited level 2 of the adjacent well. These two levels form a spatially extended doublet with the "symmetric" wavefunction $|S\rangle$ as the

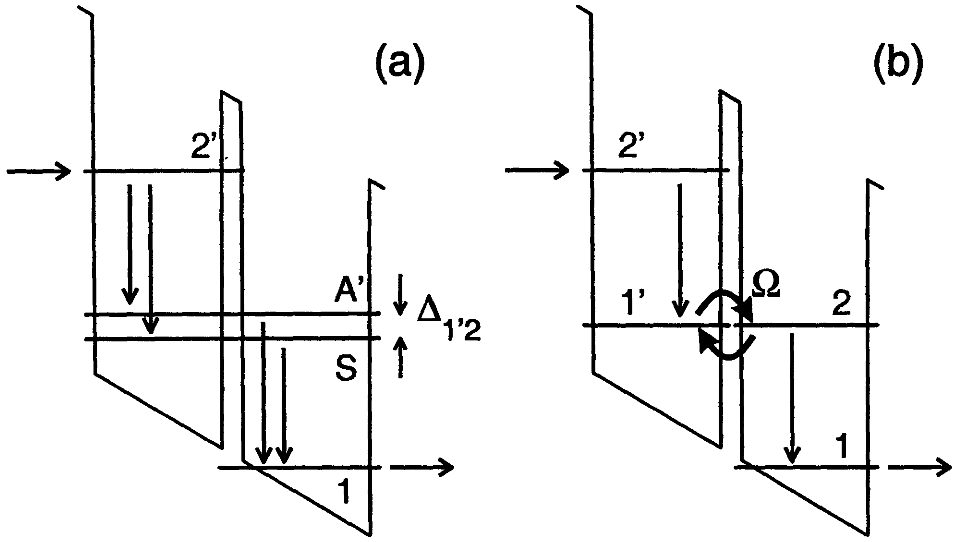

Figure 2-5: *Difference between semi-classical and coherent picture of coupled quantum wells. (a) Semi-classical picture, the wavefunctions represent eigenstates of the Hamiltonian and are delocalized at resonance. Transport through the barrier happens as soon as electrons enter levels $|S\rangle$ or $|A\rangle$. (b) Coherent picture. The wave packet is initially localized in the left well. Electrons are transported through the barrier with Rabi oscillations at frequency Ω due to the interaction between $1'$ and 2.*

lower-energy state, and the "anti-symmetric" wavefunction $|A\rangle$ as the higher-energy state (Fig. 2-5(a)). The energy separation of the doublet is the anticrossing gap $\Delta_{1'2}$. In this semi-classical picture, the transport through an energy barrier is effectively instantaneous, as both $|S\rangle$ and $|A\rangle$ are spatially extended across the barrier, and consequently the barrier causes no "resistance" to the electron transport under the resonant bias. The only "bottleneck" of this transport process is the energy-relaxing (inelastic) intersubband scattering from the doublet $|S\rangle$ and $|A\rangle$ into 1, or equivalently (in the case of a superlattice) into the doublet $|S''\rangle$ and $|A''\rangle$ formed by 1 and $2''$ of the following well (not shown in the figure). As a result, the current density under this resonant bias is independent of the barrier thickness, which is only valid in the absence of dephasing, and is thereby unphysical for real devices.

In contrast, in Fig. 2-5(b) localized basis states are used and electron transport through the barrier takes place via a coherent time evolution of these states, i.e., it takes the electrons a finite time to get from one well to the next. In this scheme, at an initial time, the electron wavepacket resides at the bottom of the left well in state 1$'$. This wavepacket can be composed as a coherent superposition of level $|S\rangle$ and level $|A\rangle$. As time evolves, this wavepacket oscillates across the middle barrier at the Rabi oscillation frequency $\Delta_{1'2}/\hbar$. In the absence of pure dephasing, this oscillation will be damped only by the intersubband scattering as the wavepacket is depleted each time it is in the right well where intersubband scattering takes place. The main bottleneck of the electron transport is again the intersubband scattering lifetime, as in the scenario of Fig. 2-5(a), even though a finite transport time (half of the Rabi oscillation period) across the barrier increases the dwell time in each well and consequently reduces the current density somewhat. Note that, in the absence of dephasing, the wave packet spends half of its time in either well, so that the time-average of the population distribution is in agreement with the picture described in Fig. 2-5(a). The most significant difference between the two schemes, however, becomes clear in the presence of dephasing scattering that may be caused by various elastic intrasubband scattering mechanisms, such as interface roughness and electron-impurity scattering. With dephasing scattering, the Rabi oscillation can be damped even in the absence of inelastic intersubband scattering. One may envision that the dephasing scattering can be so strong that the Rabi oscillation is overdamped, i.e. the time evolution of the wavepacket from the left to the right well is no longer oscillatory (which is a direct analogy to an overdamped harmonic oscillator). In this strong dephasing limit, the bottleneck of the current transport is the tunneling barrier, which is the scenario discussed by Luryi. [30] Now the time-averaged population distribution will be different in both wells, as electrons pile up behind the barrier, and the simple semi-classical picture of Fig. 2-5(a) is no longer a good approximation.

So far, most of the analysis of transport processes in QCLs has been based on the semi-classical model described in Fig. 2-5(a). [23, 12, 31, 32] This is mainly because QCLs were first developed at mid-infrared frequencies, where the photon energy $\hbar\omega > 100$ meV. Consequently, the injection barriers are relatively thin, which results in a large anticrossing gap of $\Delta_{1'2} \sim 10$ meV. Dephasing due to intrasubband scattering does not cause a significant damping to the fast $1' \leftrightarrow 2$ oscillation, and the main bottleneck of the transport is due to intersubband scattering. In fact, this Rabi oscillation at 2.5 THz ($\sim$ 10 meV) across the injection barrier has been experimentally observed in a mid-infrared QCL by using a time-resolved pump-and-probe method. [21] In THz QCLs, however, the photon energy is much smaller, $\hbar\omega \sim$ 10-20 meV. Therefore, the injection barrier must be made thicker with a smaller anticrossing gap of $\Delta_{1'2} \sim 1$ meV in order to maintain a high injection selectivity. In comparison, the dephasing rate, which can be estimated from the measured spontaneous emission linewidth, is relatively higher ($\sim$ 4-6 meV in our THz QCLs based on resonant LO-phonon scattering). [33] As a result of this much stronger dephasing relative to the injection anticrossing gap, we have found that transport analysis based on the semi-classical model is quite inadequate. For example, the measured maximum current densities at resonance are observed to be very sensitive to the thickness of the injection barriers. Also, simulations based on the semi-classical model tend to overestimate the current densities and material gains in our laser devices, [23, 12] and even predict substantial levels of gain in experimental devices that did not achieve lasing. It is this significant discrepancy between simulation results based on the semi-classical model and experimental results that motivates us to pursue the investigation described in this chapter: the importance of the coherent aspect of the transport process, or equivalently, the quantitative effect of dephasing scattering on the transport process involving subband levels at resonance.

In the following sections, we will introduce the density matrix formalism in more detail, and compare the semi-classical and density matrix approaches by investigating electron transport through a barrier in resonant and non-resonant bias conditions.

2.3.2 Density Matrix Formalism

To describe the time evolution and phase coherence of a large number of particles, we can choose from several different approaches. The most straightforward method would be to use a Schrödinger picture, where we keep track of the full wavefunction of every particle. From these wavefunctions we can then easily find the relevant macroscopic quantities, like current, population density or optical gain by summing over the contributions from each particle. However, due to the inherently statistical nature of these quantities, much of the information contained in those wavefunctions is averaged out and turns out not to be relevant for the macroscopic picture.

Another, more efficient approach is the density matrix formalism, [34] which describes the statistical distribution of quantum states in a system. This method allows us to treat the properties of a large ensemble of electrons (particles) statistically, without worrying about the exact details of the individual electrons' wavefunctions. A generic particle from this ensemble can be represented by a wavefunction ψ:

$$|\psi\rangle = \sum_i c_i |\phi_i\rangle, \tag{2.65}$$

where ϕ_i are the basis wavefunctions belonging to the Hamiltonian H_0 of the unperturbed system, and

$$c_i = \langle\phi_i|\psi\rangle. \tag{2.66}$$

We can then define the density operator

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)|, \tag{2.67}$$

which takes the form of a projection operator. The density operator can be interpreted as a description of the probability distribution in a system. For an ensemble of particles, the density matrix elements are defined as the ensemble averages:

$$\rho_{mn} =< c_n c_m^* > . \tag{2.68}$$

The diagonal elements ρ_{ii} describe the probability of finding the system in state $|i\rangle$ and are proportional to the population density of that state. The off-diagonal elements ρ_{ij} are related to the polarization between states i and j and describe the degree of coherent interaction. Consistent with its interpretation as "probability matrix," it can be shown that

$$Tr(\rho) = \sum_i^N \rho_{ii} = 1, \tag{2.69}$$

which reflects that the total population density is conserved. Another important property is that the magnitude of each off-diagonal element is smaller than or equal to the geometric mean of the corresponding diagonal elements (Schwartz inequality):

$$\rho_{ii}\rho_{jj} \geq |\rho_{ij}|^2. \tag{2.70}$$

Physically, the equality corresponds to a "pure state" described by a single wavefunction, such as the one described by Eqs. (2.65)-(2.67). Note that this pure state needs not represent just a single electron, but can instead also be used to describe time evolution of an ensemble of particles. On the other hand, the inequality in Eq. (2.70) refers to a "mixed state," which can in general be broken up into simpler constituent pure states:

$$\rho^{mixed} = \sum_i |\psi_i\rangle\langle\psi_i|. \tag{2.71}$$

A mixed state cannot be described with a single wavefunction, and represents an ensemble consisting of independently evolving pure states, i.e. a mixed state is in fact an ensemble of ensembles. In essence, a mixed state reflects the interaction between a subsystem that is characterized by a well-defined Hamiltonian H_0 and the rest of the environment, whose effects are too complex to be dealt with from first principles. The effect of dephasing can then be considered as the scrambling of the phase coherence of some electrons in one of the constituent pure state ensembles. In this picture, dephasing causes the involved electrons to be removed from their original pure state, and subsequently added back to the mixed state in a new constituent pure state, but with a phase unrelated to its original phase. The net effect is that the population remains unaffected (diagonal elements ρ_{ii}) while the average coherence ρ_{ij} decreases due to the randomization. As will be explained in more detail later, the density matrices used in the Monte Carlo simulation will generally be mixed states rather than pure states, since they describe the coherence of all electrons with the same transverse momentum $\mathbf{k}$.

To capture the dynamics of coherent transport we need the equation of motion, which describes the time-evolution of the density operator and hence of the populations and polarizations. This equation of motion is also known as the quantum Liouville equation and can be written as:

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar}[H, \rho], \tag{2.72}$$

with $H = H_0 + H'$, and H' represents a perturbation. In our case, H' consists only of an adjustment of the electron potential ΔV_{TB} due to the coupling of the localized states in one QCL module to the states in the neighboring modules, in the spirit of the tight-binding (TB) model. It is possible to rewrite the above equation so that its expression is formally identical to the calculation of a wavefunction. Looking at the right-hand side of Eq (2.72), we see that the elements of the density operator ρ_{ij} undergo a linear transformation and it is therefore possible to define a linear operator $\mathcal{L}$ to describe this transformation. The Liouville Eq. (2.72) then becomes:

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar}\mathcal{L}\rho, \tag{2.73}$$

where $\mathcal{L}$ is called the Liouville operator, and

$$\mathcal{L}_{ij,mn} = H_{im}\delta_{jn} - H^*_{jn}\delta_{im}. \tag{2.74}$$

In this representation (Liouville space), the superoperator $\mathcal{L}$ is a $N^2 \times N^2$ matrix (N is the number of states in the system) and ρ is a N^2-dimensional vector. The number of elements in $\mathcal{L}$ scales with the fourth power of N, and systems with many states can quickly pose almost insuperable computational challenges. The simplifications and approximations discussed in Section III effectively "remove" many off-diagonal elements, and vastly simplify the numerical implementation. Note that Eq. (2.73) looks very similar to the Schrödinger equation, and this allows us to apply the same formalisms to both Hilbert (wavefunction) and Liouville (density operator) spaces. Using this complete formal analogy, we can apply the calculation techniques developed for wavefunctions to the density matrix formalism and obtain the desired results.

So far we have described the coherent time evolution of an electron wavefunction in a system H_0 with only a constant perturbation ΔV_{TB} due to the interaction with the neighboring modules. Note that in the absence of scattering implicitly assumed in Eq. (2.73), the transverse momentum $\mathbf{k}$ is conserved and ΔV_{TB} is non-zero only for states with an identical $\mathbf{k}$. There is no coherent interaction between states with different in-plane momentum in this approximation. [35] As explained below, transport between states with different $\mathbf{k}$ is handled separately, through semi-classical scattering mechanisms within the same module. A fully coherent description of the many interactions, such as electron-phonon, electron-impurity and electron-electron scattering, would be very involved and computationally intensive. Therefore, it is more convenient to describe them as semi-classical scattering events, in which electron scattering rates are described by Fermi's Golden Rule. In view of the setup of our model, this is a reasonable approximation. In THz QCLs, the injection barrier is usually much thicker than the other (intramodule) barriers, which results in the "intermodule" anticrossing gaps (where each level is on a different side of the injection barrier) being considerably smaller than "intramodule" anticrossing gaps. As explained in greater detail in Section IV, the intermodule interactions are thus more sensitive to dephasing, and are best described with Eq. 2.73, while for the intramodule transitions a semi-classical model is adequate. The scattering events affect the transport of electrons between different states within the system, and hence cause the relaxation of the population with a scattering time T_1. In addition to this relaxation scattering, we can also consider "pure dephasing" events that merely scramble the phase correlation between two states at a rate T_2^{-1} without causing depopulation. As mentioned before, pure dephasing accounts for the effects of the bath on the electrons in the system, and as such describe scattering events that are not explicitly included in the simulation model. Both relaxation scattering and pure dephasing contribute to the dephasing time τ_{deph}:

$$\frac{1}{\tau_{\text{deph}}} = \frac{1}{2T_1} + \frac{1}{T_2}. \tag{2.75}$$

We note that the contribution of the relaxation scattering is half of that of the pure dephasing. This is due to the fact that T_2 describes the relaxation of the polarization ρ_{ij}, which is proportional to the amplitude of the oscillation ($\rho_{ij} \propto e^{-t/T_2}$); on the other hand, T_1 is a probability decay rate that reduces $\rho_{ii} \propto |\rho_{ij}|^2$, proportional to the energy density ($\rho_{ii} \propto e^{-t/T_1} \rightarrow \rho_{ij} \propto e^{-t/2T_1}$). In the equations of motion, scattering and pure dephasing add extra relaxation terms to the expressions for ρ_{ii} and ρ_{ij}, which can be incorporated in the density matrix formalism with a corresponding superoperator $\mathcal{F}$. Eq. (2.73) now becomes:

$$\frac{\partial \rho_{ij}}{\partial t} = \sum_{m,n} (\mathcal{L}_{ij,mn} + \mathcal{F}_{ij,mn})\rho_{mn}, \tag{2.76}$$

with

$$\mathcal{F}_{ij,mn} = -\frac{1}{2}(\gamma_i + \gamma_j)\delta_{im}\delta_{jn} - \Gamma_{ij}^{\text{pure}}\delta_{im}\delta_{jn}, i \neq j, \tag{2.77}$$

and

$$\mathcal{F}_{ii,jj} = \gamma_{ji}(1 - \delta_{ij}) - \gamma_i\delta_{ij}. \tag{2.78}$$

Here γ_i and γ_j correspond to the total scattering rates out of the i and j levels, γ_{ji} is the net scattering rate from level j to level i and $\Gamma_{ij}^{\text{pure}}$ ($=T_2^{-1}$) is the pure dephasing rate of the coherent transport between i and j. It is important to note that the dephasing of ρ_{ij} is not due only to scattering between i and j, but rather to *all* scattering events involving either i or j. Electrons scattering out of a level i disrupt the coherent transport from and to i, and as such dephase all polarizations ρ_{ik} ($k \neq i$) that involve level i. In Eq. (2.76), all the stochastic aspects of the electron transport are included in the operator $\mathcal{F}$. Using a phenomenological pure dephasing time constant T_2, Eq. (2.76) can be solved analytically to yield an expression of current density. [36, 37]

2.3.3 Superlattice

To quantitatively illustrate the effects of dephasing and coherence on transport through a barrier, it is instructive to investigate a simple superlattice structure as sketched in Fig. 2-6(a), consisting of a succession of coupled quantum wells, separated by a barrier with thickness t_{barr}, and with two energy levels 1 and 2 in each well. Every

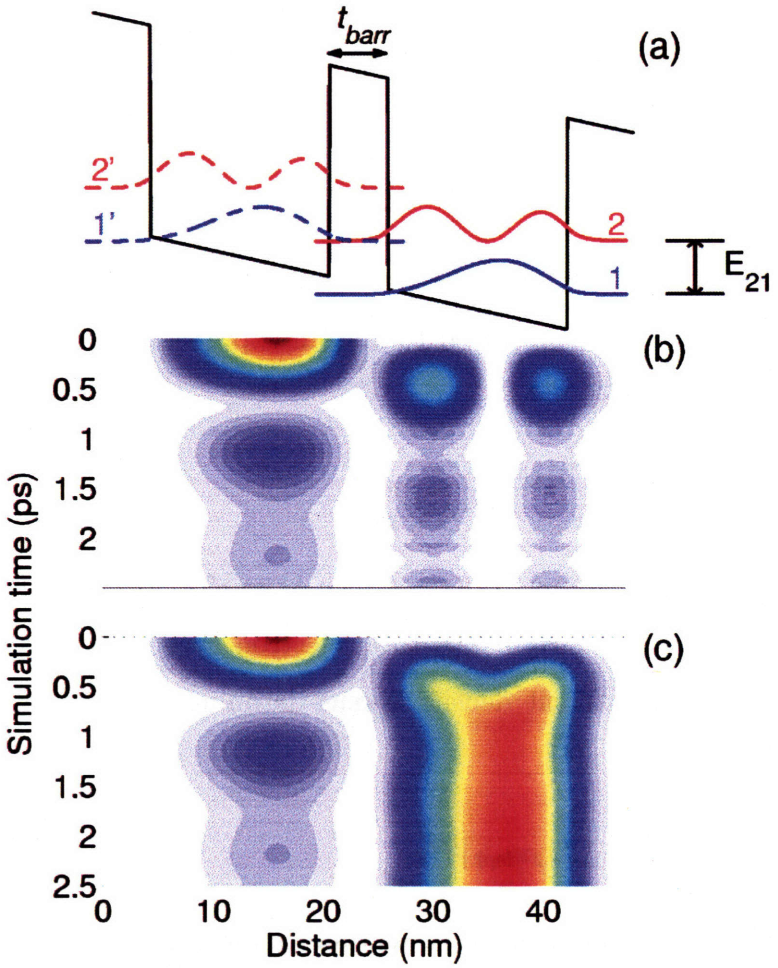

Figure 2-6: *Time evolution of the population density in a superlattice with 16.8-nm-wide GaAs wells and 4.1-nm-wide* $Al_{0.15}Ga_{0.85}As$ *barriers. The ground level in one well is in resonance with the first excited level in the next well* ($\Delta_{1'2} = 3.9$ *meV*)*. Level* $n = 2$ *is depopulated by LO-phonon scattering* ($\tau_{21}^{\mathrm{LO}} \approx 0.3$ *ps*)*. A population density of* 1×10^{10} cm^{-2} *is initially placed in* $n = 1'$*. (a) Conduction band profile and wavefunction probability distributions of the generic superlattice structure used throughout this paper. Also indicated are the barrier thickness* t_{barr} *and energy separation* E_{21}*, which are referenced throughout the text. (b) Spatial distribution of the populations of* $n = 1'$ *and* $n = 2$ *versus time. The damped Rabi oscillations of the population density are clearly visible. (c) Time evolution of the total population density in* $n = 1'$, 2, *and the ground state* 1.

well corresponds to a module, and the wavefunctions are localized in the wells using the procedure outlined above. The anticrossing gap between $n = 1'$ and $n = 2$ is given by $\Delta_{1'2} = 2\Delta V_{1'2}$. The equations of motion can be written as: [38]

$$\frac{d}{dt}(\rho_{11} - \rho_{22}) = \frac{2i\Delta V_{1'2}}{\hbar}(\rho_{21} - \rho_{21}^*) - \frac{(\rho_{11} - \rho_{22}) - (\rho_{11} - \rho_{22})_0}{\tau_2}, \tag{2.79}$$

$$\frac{d}{dt}\rho_{21} = \frac{i\Delta V_{1'2}}{\hbar}(\rho_{11} - \rho_{22}) - \frac{iE_{1'2}}{\hbar} - \frac{\rho_{21}}{\tau_{\text{deph}}}, \tag{2.80}$$

where $(\rho_{11}-\rho_{22})_0$ is the population difference at equilibrium, and $\tau_{\text{deph}}^{-1} = 0.5\tau_2^{-1}+T_2^{-1}$. These coupled equations can be solved to find an expression for the current density through the barrier: [36, 37]

$$J = qN_s\frac{|\Omega|^2\tau_{\text{deph}}}{1 + (E_{1'2}/\hbar)^2\tau_{\text{deph}}^2 + \Omega^2\tau_2\tau_{\text{deph}}}, \tag{2.81}$$

where $E_{1'2}$ is the energy detuning from resonance, $\Omega = 2\Delta V_{1'2}/\hbar$ is the Rabi oscillation frequency at resonance, and N_s is the total electron sheet density per well. This expression describes the current density versus detuning bias $E_{1'2}$ as a Lorentzian with a full-width half-max of

$$\Delta E_{\text{FWHM}} = \frac{2\hbar}{\tau_{\text{deph}}}[1 + \Omega^2\tau_2\tau_{\text{deph}}]^{1/2}. \tag{2.82}$$

The density matrix model provides a picture of a wavepacket oscillating between $n = 1'$ and $n = 2$ at a frequency Ω with a damping rate of τ_{deph}^{-1}. This is illustrated in Fig. 2-6(b) and (c), which show the spatial distribution of the population density in a superlattice designed to study the dynamics of the depopulation of the lower radiative level 4 in FL177C-M5. The electron extraction from level 4 critically relies on the anticrossing with another level 3, which is subject to resonant LO-phonon scattering ($\tau_{\text{LO}} < 0.3$ ps). In our simple model we identify 4 and 3 (in the QCL) with the ground state $1'$ and the first excited state 2, respectively. The anticrossing energy $\Delta_{1'2} = 2\Delta V_{1'2}$ is 3.9 meV and E_{21} is 39 meV. A population of 1×10^{10} cm^{-2} is initially situated in $1'$ and then oscillates back and forth between the anticrossed levels $1'$ and 2. There is no pure dephasing added in this simulation ($1/T_2$=0). Electrons in state 2 are subject to resonant LO-phonon depopulation, which results in a damped

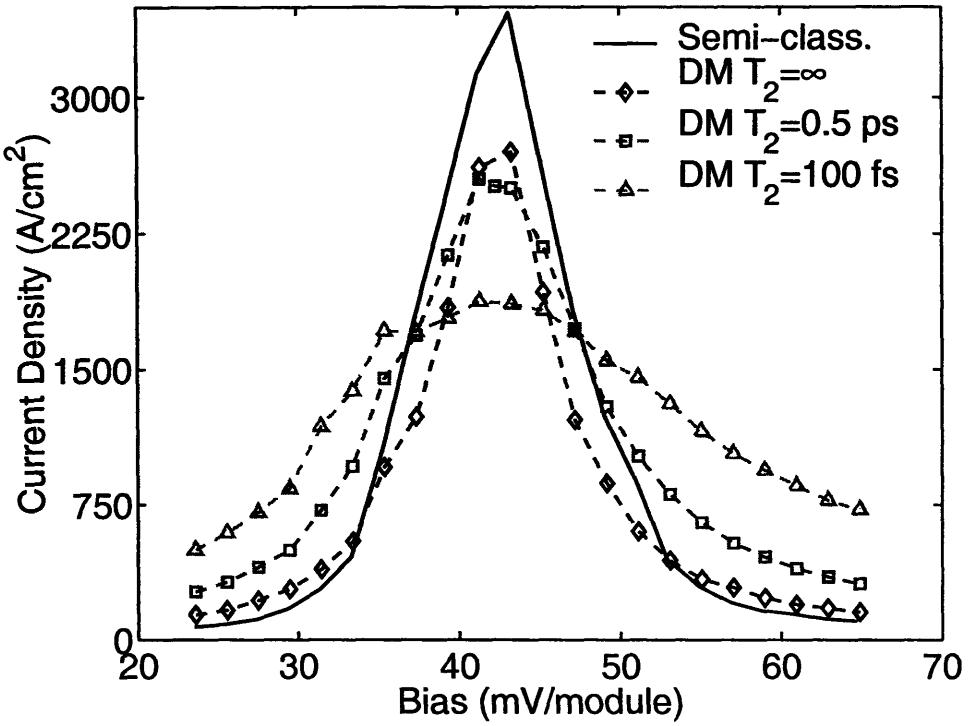

Figure 2-7: *Dependence of current density on bias for a superlattice consisting of 15.5-nm-wide GaAs wells, separated by 4.1-nm* $Al_{0.15}Ga_{0.85}As$ *barriers. The energy separation* $E_{21} \approx 42$ *meV and* $\Delta_{1'2}$*=4.5 meV. Results are shown for both the semi-classical and the density-matrix simulations.*

Rabi oscillation with $\tau_{\text{deph}} \approx 0.6$ ps. From Fig. 2-6(c), we can see it takes about 1.5 ps for the majority of the electrons to transfer from 1′ (corresponding to the lower radiative level 4) to 1, which corresponds to the relaxation/injection states in the QCL. This is slower than what we would expect from the semi-classical picture, where $\tau_4 \approx 0.5$ ps. [11] Note that even in the absence of dephasing, it would take approximately $\hbar/2\Omega \approx 0.5$ ps for the electrons to oscillate across the barrier, which largely explains the longer dwell time in the density matrix calculation.

The dependence of the current density on the pure dephasing rate T_2^{-1} is shown in Fig. 2-7 for a similar superlattice. At resonance, the peak current density decreases with T_2^{-1} as the Rabi oscillation is increasingly damped and it becomes more and more difficult for electrons to tunnel through the barrier. The difference between the semi-classical result and the density matrix simulation in the limit of T_2^{-1}=0

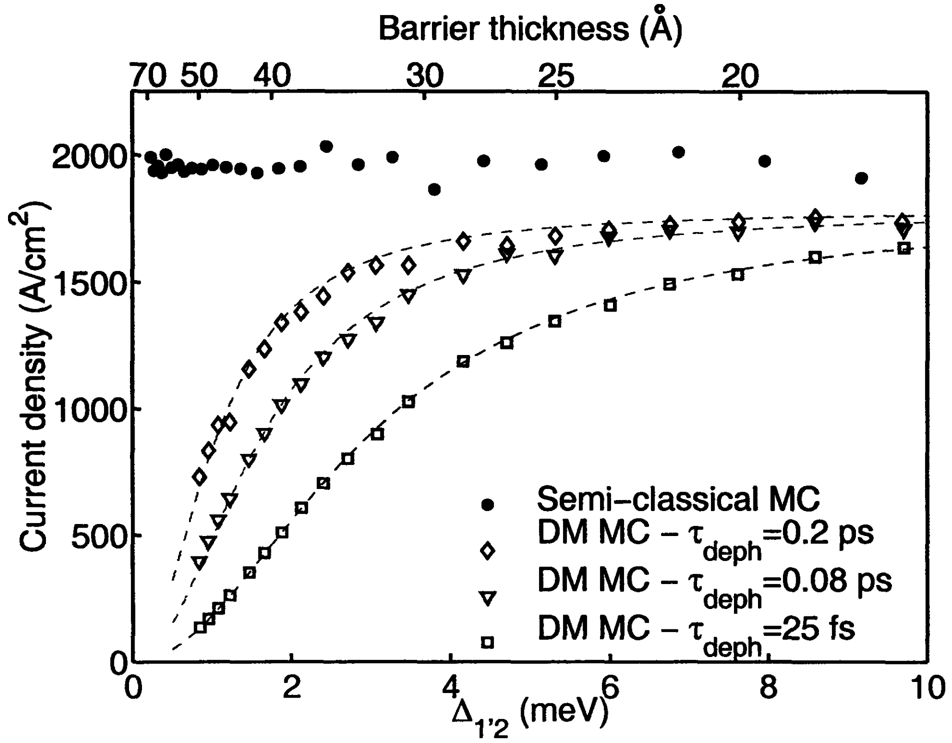

Figure 2-8: *Dependence of peak current density on anticrossing gap* $\Delta_{1'2}$ *for a* $GaAs/Al_{0.3}Ga_{0.7}As$ *superlattice with 14.8-nm thick wells separated by barriers with varying thicknesses. The energy separation* $E_{21} \approx 60$ *meV. The semi-classical results are independent of* $\Delta_{1'2}$ *and the barrier thickness, while the density matrix simulation clearly shows the decrease of the peak current density with smaller* $\Delta_{1'2}$ *and higher dephasing rate* T_2^{-1}. *The dashed lines represent theoretical calculations using Eq. (2.81).*

is due to the relaxation dephasing, which is the term $-1/2(\gamma_i + \gamma_j)$ in Eq. (2.77). Away from resonance, the DM model predicts a broader $I - V$ curve with a higher current density than the semi-classical picture. This is due to the level broadening (Eq. (2.82)) which relaxes the energy alignment of $n = 1'$ and $n = 2$. Especially at high dephasing rates, the increased interaction can cause the upper radiative state (not shown here) to couple more strongly to the injector/relaxation level 1, and thus reduce the depopulation selectivity.

The influence of the barrier thickness t_{barr} is felt through a diminished coupling $\Delta V_{1'2}$, and can be illustrated by investigating the current transport through a barrier. We consider a superlattice structure similar to the one shown in Fig. 2-6(a), with an energy separation $E_{21} \approx 60$ meV which is much larger than $\hbar\omega_{\mathrm{LO}}$. This ensures that, in the semi-classical picture, the intrawell energy difference E_{21} still exceeds $\hbar\omega_{\mathrm{LO}}$ even for large anticrossing gaps so that scattering from $n = 2$ is dominated by LO-phonon scattering ($\tau_2^{\mathrm{LO}} \approx 0.3$ ps, and nearly independent of E_{21}). The peak current density occurs when the lower level $1'$ of one well lines up with the upper level 2 of the adjacent well. In the semi-classical approach, the anticrossed wavefunctions are delocalized across both wells, and share an identical carrier lifetime τ (independent of barrier thickness) disregarding the minor energy shift due to the anticrossing. This leads to the unphysical result that the peak current density in the semi-classical description does not depend on the barrier thickness (as shown in Fig. 2-8). In other words, in this picture electrons scatter from one spatially extended state into the next, and never experience any effect from the barrier whatsoever, as illustrated previously in Fig. 2-5. To include tunneling effects, and hence more accurately describe transport through a barrier, we need to take into account the phase correlation between localized basis states, so that the result of phase relaxation is a collapse of the electron wavefunction into localized states and an interruption of resonant tunneling.

The results for the density matrix calculations are also shown in Fig. 2-8. In contrast to the semi-classical results, the DM-MC calculations reveal a strong dependence of the peak current density on the anticrossing gap (or barrier thickness) and the pure dephasing time T_2. For large anticrossing gaps or thinner barriers, the Rabi oscillation frequency is much higher than the dephasing rate and the peak current density approaches the semi-classical limit. In this regime the population is spread equally across both subbands, and the current density is given approximately by $J = qN_s/2\tau_2$. For smaller anticrossing gaps, or thicker barriers, the dephasing scattering becomes increasingly important and inhibits the transport through the barrier, which results in a lower peak current density. From Eq. (2.81) we can see that the current density starts to roll off when $\Delta_{1'2} \approx \hbar/\sqrt{\tau_2\tau_{\mathrm{deph}}}$. The dashed lines are analytical results calculated from Eq. (2.81), which agree well with our DM-MC results. The good agreement based on this simple structure gives us confidence to investigate more complicated structures using the numerical DM-MC tool developed in this work.

Note that localized basis wavefunctions are necessary to produce the above result within the density matrix formalism. If we chose spatially extended basis wavefunctions (as in the semi-classical case), a localized wavepacket would be represented as a coherent superposition of extended wavefunctions:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|S\rangle + |A\rangle) \;\rightarrow\; \rho = |\psi\rangle\langle\psi|. \tag{2.83}$$

In the absence of dephasing (and scattering), the time evolution of this wavepacket would be identical to the one described with localized basis wavefunctions. A proper implemention of scattering requires a full density matrix approach, [39] which is not expected to result in a substantial difference between the two models. However, the inclusion of pure dephasing causes the off-diagonal elements of ρ to decrease, so that even in the strong dephasing limit of $T_2 \rightarrow 0$:

$$\rho = \frac{1}{2}|S\rangle\langle S| + \frac{1}{2}|A\rangle\langle A|, \tag{2.84}$$

which is still equivalent to the semi-classical model, and no decrease in current density is predicted. This means that extended wavefunctions are not a good choice for the basis wavefunctions in our model.

Chapter 3

Monte Carlo Simulation

3.1 Boltzmann Transport Equation

In a semi-classical picture, the movement of carriers in real space **r**, momentum space **k** and time is fully described by the Boltzmann Transport Equation (BTE). For a system with a constant number of electrons (such as the QC structures we are interested in) this equation of motion in the presence of an in-plane electric field $\mathcal{E}$ is written as:

$$\frac{\partial f_i}{\partial t} + \mathbf{v} \cdot \nabla_r f_i + \mathcal{E} \cdot \frac{1}{\hbar} \nabla_k f_i = \left. \frac{\partial f_i}{\partial t} \right|_{\text{coll}}, \tag{3.1}$$

where $\mathbf{v}$ is the in-plane electron velocity, f_i is the electron distribution function of subband i, and

$$\nabla_r f_i \equiv \frac{\partial f_i}{\partial x} \hat{\mathbf{x}} + \frac{\partial f_i}{\partial y} \hat{\mathbf{y}} \tag{3.2}$$

and

$$\nabla_k f_i \equiv \frac{\partial f_i}{\partial k_x} \hat{\mathbf{x}} + \frac{\partial f_i}{\partial k_y} \hat{\mathbf{y}} \tag{3.3}$$

denote the divergence of f_i in real and momentum space, respectively. Remark that the last two terms on the left-hand side of the BTE only describes the in-plane motion of the electron, which is of little interest for us. Instead, the collision term on the right hand side includes both intra- and intersubband scattering and its calculation and implementation lie at the core of our transport calculation.

The Boltzmann Transport Equation is very hard to solve analytically; it is much easier to follow the trajectories of selected individual electrons as they move through a device under the influence of electric fields and scattering forces. Each of these paths is random as far as the occurrence of particular scattering events is concerned, If we follow a number of electrons, sufficient to describe the distributions in the different subbands of a structure, the averaged results will be a good approximation of the average behavior of the electrons within the device. The importance of random events (and random numbers) in this type of calculation has given rise to the name "Monte Carlo" (MC) simulation. In many cases, MC simulation is the most accurate technique available for simulating transport in devices, and it is considered the standard against which the validity of simpler approaches is measured.

In the following sections we will take a closer look at the underlying principles of the Monte Carlo simulation, and at its practical implementation. Because it directly mirrors the physics, a better understanding of the technique will also allow for a better intuitive picture of the simulation results.

3.1.1 Free Flight

Typically, the motion of an electron in a device is due to two contributions: the presence of on electric or magnetic field, accelerating the electron and changing its energy continuously, and the occurrence of scattering events. Because the latter are usually much shorter in duration than the "free flight" in between collisions, scattering is assumed to be instantaneous and change the electron's momentum and energy abruptly.

To simulate a typical single free flight and scattering event, four different random numbers are used. The first specifies the duration of the free flight, during which the electron moves in accordance with Newton's laws. At the end of the free flight, the electron's momentum and energy are updated. The next random number determines the type of scattering event, and two more numbers are needed to determine the electron's final state, which is characterized by its momentum amplitude and direction. Once the final state is known, a new free flight is initiated.

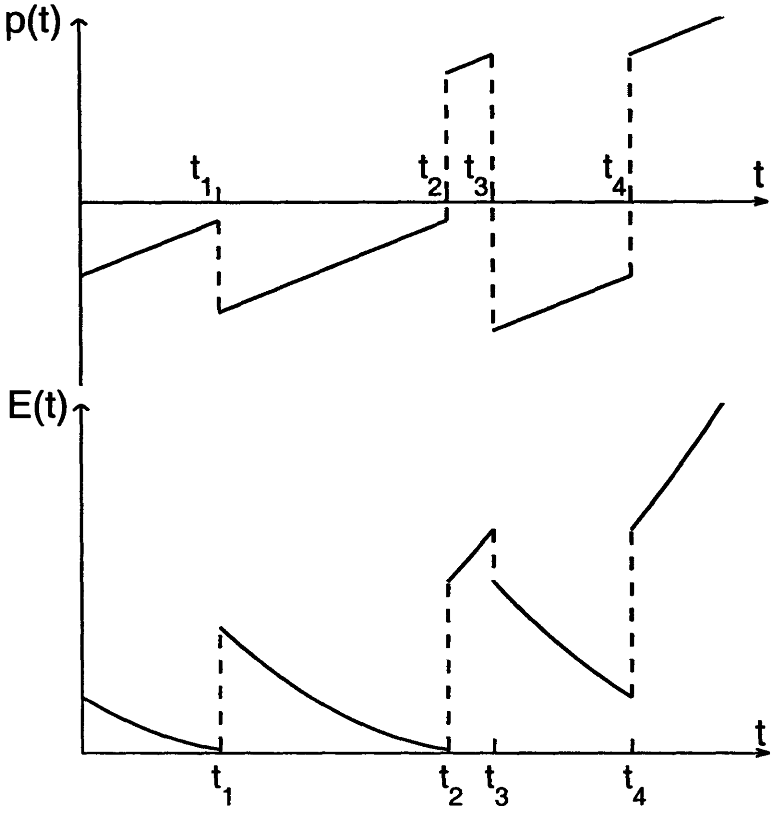

Figure 3-1: *Time evolution of momentum and energy of a particle. Scattering events occur at times* $t_1, ..., t_4$*, separated by periods of free flight during which the momentum and energy of the electron change continuously.*

In QC structures, an electron is completely characterized by its subband i and its in-plane momentum $\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}}$, where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the unit vectors of the in-plane directions. The total energy can then be derived from the subband edge energy E_i^0 and the electron kinetic energy $E^k = \hbar^2k^2/2m^*$. Note that the momentum in the growth direction, k_z, is not a good quantum number (due to the localization of the wavefunctions in the z-direction) and has a distribution that corresponds to the Fourier spectrum of the electron's wavefunction. As such, the effect of electric fields on k_z is felt through their effect on the wavefunction, and cannot be described by a simple relation.

If we apply an electric field $\mathcal{E}_x$ in the x-direction (in-plane), after a time t we find the momentum $\mathbf{k}(t)$ and energy $E(t)$ of an electron from:

$$k_x(t) = k_x(0) + \frac{-q}{\hbar}\mathcal{E}_x t$$

$$k_y(t) = k_y(0) \tag{3.4}$$

and

$$E(t) = E_i^0 + \frac{\hbar^2}{2m^*}[k_x^2(t) + k_y^2(t)] \tag{3.5}$$

The duration of the free flight is determined by the total scattering rate $\Gamma_i(\mathbf{k})$ of the followed electron, which is the sum of the scattering rates $1/\tau_i(\mathbf{k})$ of the different scattering mechanisms, such as e-LO-phonon, e-e and e-imp scattering:

$$\Gamma_i(\mathbf{k}) = \frac{1}{\tau_i^{\text{e-LO}}(\mathbf{k})} + \frac{1}{\tau_i^{\text{e-e}}(\mathbf{k})} + \frac{1}{\tau_i^{\text{e-imp}}(\mathbf{k})} + \cdots \tag{3.6}$$

As the electron's subband and momentum change during its time evolution, in general Γ is also time-dependent and the free-flight time will change over the course of the simulation. However, for now we can simplify the model and assume a time-independent scattering rate Γ_0. As we will see, the conclusions we reach are still general and applicable even for a time-dependent Γ.

How do we properly choose the free flight time t_f so that it reflects the random nature of the scattering mechanisms? To address this problem, we can investigate the time evolution of a population n_i of electrons in a subband i, subject to a constant scattering rate $\Gamma_0 = 1/\tau$. In other words, during an infinitesimal time interval dt, each electron in i has a constant probability $\Gamma_0\, dt$ to undergo a collision and be removed from the subband:

$$dn_i = -\Gamma_0 n_i\, dt, \tag{3.7}$$

which can be solved to find:

$$n_i(t) = n_i(0)e^{-\Gamma_0 t}. \tag{3.8}$$

The population of the subband declines exponentially, and the probability that one selected electron will survive until time t decreases correspondingly with $e^{-\Gamma_0 t}$. The probability $\mathcal{P}(t)$ that an electron will scatter between times t and $t + dt$ depends on the scattering rate and the chance that the electron is still present in i:

$$\mathcal{P}(t)\, dt = \Gamma_0\, e^{-\Gamma_0 t}\, dt. \tag{3.9}$$

The eventual value for $\mathcal{P}$ at the time when an electron scatters is a random value between $t = 0$ ($\mathcal{P}(0)$=0) and $t = \infty$ ($\mathcal{P}(\infty)$=1), which is unknown at the beginning of the free flight. Therefore, picking random numbers for $\mathcal{P}$ will yield a distribution of free flight times t_f that reflect the random nature of the scattering events:

$$t_f = -\frac{1}{\Gamma_0} \ln(\mathcal{P}). \tag{3.10}$$

As mentioned earlier, in reality the scattering rate is not constant but varies with the subband index and momentum of the electron. The difference between our proposed constant Γ_0 and the real rate $\Gamma_i(\mathbf{k})$ is referred to as the "self-scattering" rate $\Gamma_i^{\text{self}}(\mathbf{k})$:

$$\Gamma_i^{\text{self}}(\mathbf{k}) = \Gamma_0 - \Gamma_i(\mathbf{k}). \tag{3.11}$$

Note that the self-scattering rate depends on the electron's momentum and subband, and that the chosen Γ_0 must be greater than the maximum of the $\Gamma_i^{\text{self}}(\mathbf{k})$ it represents. To fit in with Eq. 3.6, we can describe a self-scattering event as a fictitious scattering event, in which the electron remains unaffected and its subband and momentum stay constant. Its only reason for existing is to allow for a constant total scattering rate, which in turn considerably simplifies the calculations. With the addition of self-scattering, the total scattering rate is now constant, and we can apply Eq. 3.10 to find the free flight durations. When a self-scattering event terminates the free flight, the momentum of the electron is updated according to Eq. 3.4, and a new random number is generated to determine the duration of the next free flight. In Figure 3-2 the momentum and energy evolution of an electron are illustrated in the presence of self-scattering. It can be seen that the various fictitious scattering events do not affect the electron's trajectory. However, the inclusion of self-scattering can lead to a significant "oversampling" of the trajectory between real scattering events, and an

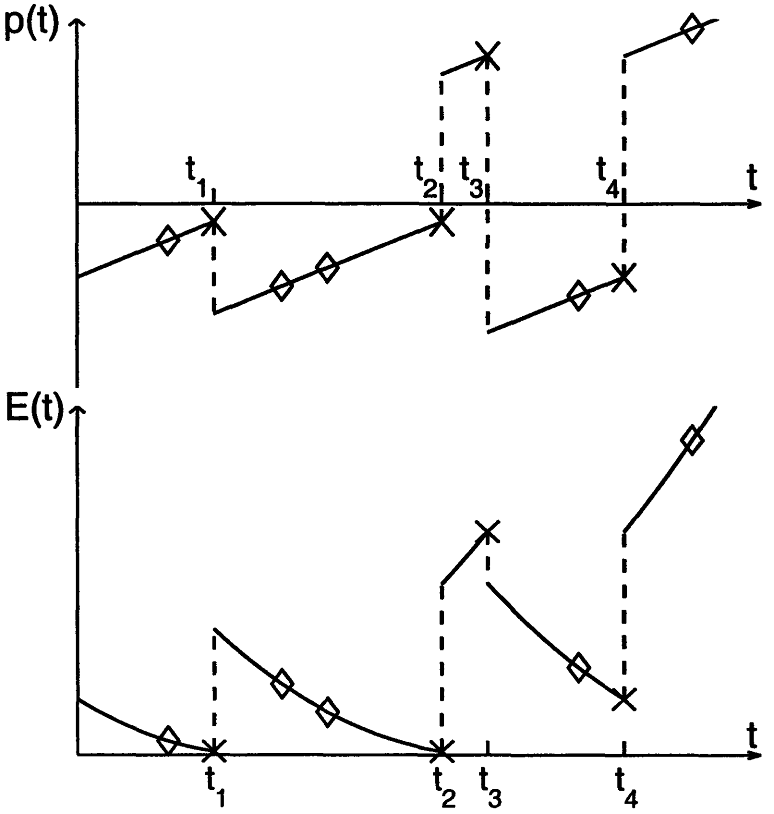

Figure 3-2: *Simulation of the time evolution of the momentum and energy of an electron under the influence of an in-plane electric field and scattering. The simulated scattering events are indicated with markers; diamonds indicate self-scattering events, crosses indicate true scattering events.*

added computational load. Therefore, it is best to choose Γ_0 close or equal to the maximum value of $\Gamma_i(\mathbf{k})$.

After a free flight is concluded, the electron's momentum and energy are updated accordingly. The next task is to select a scattering mechanism, and find how it in turn affects the electron. To do this, we need to take a closer look at the different contributions to Γ_0.

We previously mentioned that a good value for Γ_0 would be the maximum scattering rate of an electron in a system, as this minimizes self-scattering. On the other hand, this choice of Γ_0 makes it harder to efficiently select a scattering mechanism for any random electron. Indeed, the importance of the different scattering mechanisms for the electron with the maximum total scattering rate is likely to be different from

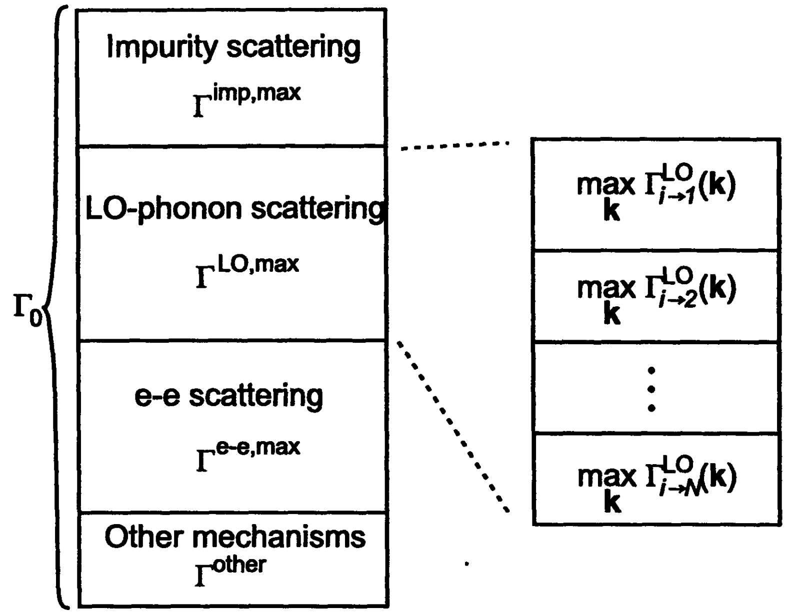

Figure 3-3: *Construction of* Γ_0 *from its constituent scattering rates.*

that for other electrons. This means that, although we know the maximum scattering rate, we have no further information about what scattering mechanism is more or less likely to be important without carrying out the relevant calculations for every electron we describe.

Therefore, it is more practical to define Γ_0 as the sum of the maximum possible scattering rates $\Gamma^{\max}$ of every contributing scattering mechanism:

$$\Gamma_0 = \Gamma^{\text{e-LO,max}} + \Gamma^{\text{e-e,max}} + \Gamma^{\text{e-imp,max}} + \cdots \tag{3.12}$$

These maximum scattering rates $\Gamma^{\max}$ are found by summing, for every initial subband i, the maximum scattering rates to all possible final subbands f, and then picking the highest resulting rate:

$$\Gamma^{\text{e-e/LO/imp,max}} = \max_j \Gamma_j^{\text{e-e/LO/imp,max}} = \max_j \sum_f \max_{\mathbf{k}} \Gamma_{j\to f}^{\text{e-e/LO/imp}}(\mathbf{k}). \tag{3.13}$$

Figure 3-3 illustrates how Γ_0 is found from the individual scattering rates. In general, this choice of Γ_0 greatly overestimates the actual scattering rate for any electron in the system, and leads to much self-scattering. However, as will be explained next, this choice of Γ_0 greatly simplifies the Monte Carlo process to determine a suitable final state when a scattering event occurs.

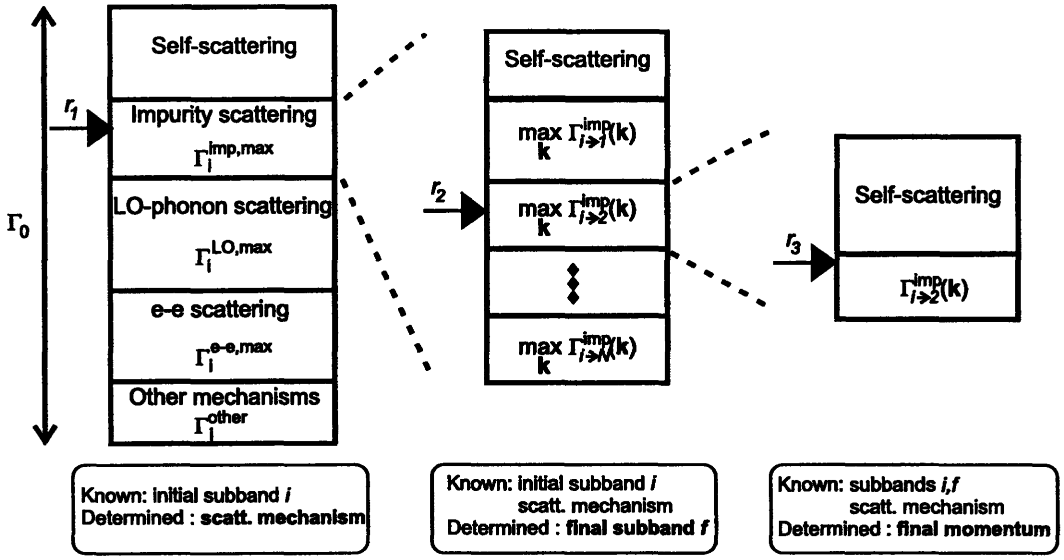

Figure 3-4: *Graphical representation of the selection process of a scattering mechansim, and the final subband and momentum by a Monte Carlo procedure.*

Figure 3-4 schematically shows how the selection process of the final state unfolds. When the momentum of the electron (with initial subband i) is updated after a free flight, a random number r_1 between 0 and Γ_0 is generated and mapped to the (maximum) scattering rates belonging to different scattering mechanisms. If we have m scattering mechanisms (labeled $j = 1, ..., m$) with rates $\Gamma_i^{\mathrm{j,max}}$, mechanism l will be selected if

$$\sum_{j=1}^{l-1} \Gamma_i^{\mathrm{j,max}} \le r_1 < \sum_{i=1}^{l} \Gamma_i^{\mathrm{j,max}}. \tag{3.14}$$

When the scattering mechanism is identified and not fictitious (self-scattering), the final subband f can be determined by a very similar procedure; a new random number r_2 $(0 < r_2 < \Gamma_i^{\mathrm{l,max}})$ is selected and the final subband f is found from:

$$\sum_{j=1}^{f-1} \Gamma_{i\to j}^{\mathrm{j,max}} \le r_2 < \sum_{i=1}^{l} \Gamma_{i\to j}^{\mathrm{j,max}}. \tag{3.15}$$

Depending on the specific scattering mechanism, the determination of the final momentum may require additional random numbers. In the following sections, we will discuss in some detail how this process works for e-LO-phonon, e-imp and e-e scattering.

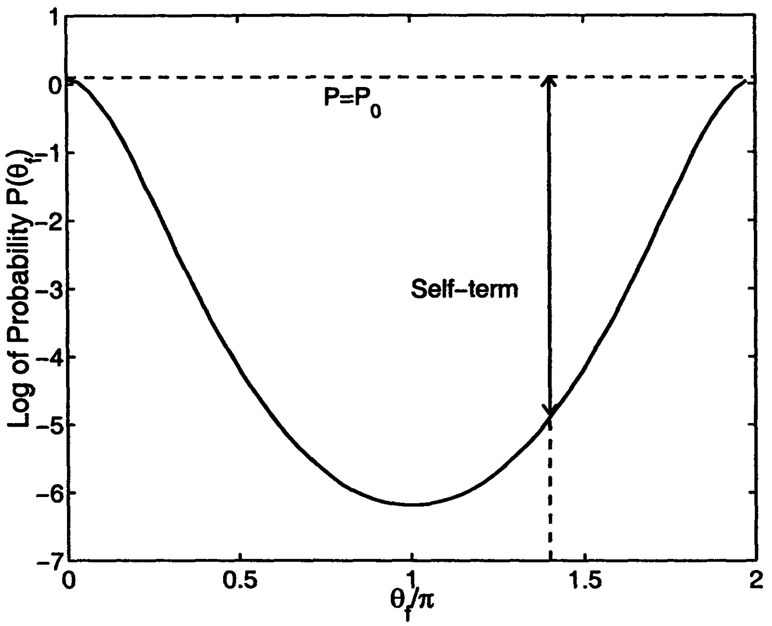

Figure 3-5: *Probability function $\mathcal{P}$ as a function of the angle θ_f. Also indicated is the maximum probability P_0.*

3.1.2 Impurity, interface roughness and LO-phonon scattering

Once the scattering process and the final subband are identified, all that remains to be determined is the final momentum $\mathbf{k}_f = k_f(\cos\theta_f\hat{\mathbf{x}} + \sin\theta_f\hat{\mathbf{y}})$. The amplitude k_f can easily be found from the conservation of energy:

$$\frac{\hbar^2 k_f^2}{2m^*} = E_i^0 - E_f^0 + \frac{\hbar^2 k_i^2}{2m^*}. \tag{3.16}$$

To determine the angle θ_f, we need to examine Eq. 2.58 more closely. We see that the total scattering rate is found by integrating over θ_f, or equivalently, that the probability $\mathcal{P}(\theta_f)\,d\theta$ of having a final angle between θ_f and $\theta_f + d\theta$ is given by:

$$\mathcal{P}^{\mathrm{imp}}(\theta_f) = \frac{\left|\frac{A_{if}^{\mathrm{imp}}(q)}{q}\right|^2}{\int_0^{2\pi} d\theta_f' \left|\frac{A_{if}^{\mathrm{imp}}(q)}{q}\right|^2} \tag{3.17}$$

where the exchanged momentum q is a function of θ_f. As illustrated in Fig. 3-5, we can find a suitable distribution of θ_f by using another, simple Monte Carlo calculation.

We can initially approximate $\mathcal{P}$ as a uniform distribution $\mathcal{P}(\theta_f) = \mathcal{P}_0$ by adding a "self-term" to $\mathcal{P}(\theta_f)$. Assuming this uniform distribution, all θ_f are equally likely and we can pick θ_f randomly between 0 and 2π. As a final step, the self-term is taken into account by generating a random number r $(0 < r < 1)$ and rejecting the choice of θ_f if

$$r > \frac{\mathcal{P}(\theta_f)}{\mathcal{P}_0}. \tag{3.18}$$

Interface roughness scattering, another elastic scattering process, can be treated in a very similar way. The handling of LO-phonon scattering, both emission and absorption, differs in the calculation of k_f:

$$\frac{\hbar^2 k_f^2}{2m^*} = E_i^0 - E_f^0 + \frac{\hbar^2 k_i^2}{2m^*} \pm \hbar\omega_{\mathrm{LO}}, \tag{3.19}$$

where the plus sign corresponds to LO-phonon absorption and the minus sign to LO-phonon emission. The explicit form of $\mathcal{P}(\theta_f)$ is:

$$\mathcal{P}^{\mathrm{LO}}(\theta_f) = \frac{\frac{[F_{if}^{\mathrm{LO}}(q)]}{q}}{\int_0^{2\pi} d\theta'_f \, \frac{[F_{if}^{\mathrm{LO}}(q)]}{q}}. \tag{3.20}$$

3.1.3 Electron-electron scattering

For e-e scattering, many more parameters need to be determined: the final subband f and momentum $\mathbf{k}_f$ of the followed electron, and the identity of the partner electron (initial subband j and momentum $\mathbf{k}_j$) and its final subband g and momentum $\mathbf{k}_g$. Because a full calculation of all involved scattering times is very time consuming and depends on the specific electron distributions, e-e scattering is handled slightly differently from the single-electron scattering processes discussed in the previous section.

Since we need to have access to information about the scattering rates for different initial and final subbands for both the followed electron and the partner electron, it is useful to define a few additional "lumped" scattering rates to describe the maximum scattering rate involving certain combinations of initial and final subbands for the electrons involved. We will model these scattering rates so that they are most helpful to us when we are trying to use MC methods to determine the various subbands involved in an e-e scattering event. It is easiest and most intuitive to start from a specific scattering event, in which the followed electron scatters from subband i to f and the partner electron from j to g. For this event we can define a maximum scattering rate $\Gamma_{ij\to fg}^{\text{e-e,max}}$ out of $|i,\mathbf{k}\rangle$ due to the described e-e scattering process:

$$\Gamma_{ij\to fg}^{\text{e-e,max}} = \max\nolimits_{\mathbf{k}} W_{ij\to fg}^{\text{e-e,max}}(\mathbf{k}). \tag{3.21}$$

This maximum scattering rate can be found by maximizing the integrand in Eq. 2.47, i.e. by taking the maximum value of $|A_{ijfg}(q)|$ over all q:

$$\frac{|A_{ijfg}(q)|^2}{|q+q_{\text{scr}}|^2} < \frac{|A_{ijfg}^{\max}|^2}{q_{\text{scr}}^2}. \tag{3.22}$$

After some manipulation, we find:

$$\Gamma_{ij\to fg}^{\text{e-e,max}} = \frac{\pi e^4}{2\hbar(4\pi\epsilon)^2}\frac{|A_{ijfg}^{\max}|^2}{q_{\text{scr}}^2} n_j \tag{3.23}$$

where n_j is the population density in subband j.

If the partner electron's final subband is unknown, the good estimate for the total scattering rate is given by the summing Eq. 3.21 over g:

$$\Gamma_{ij\to f}^{\text{e-e,max}} = \sum_g \Gamma_{ij\to fg}^{\text{e-e,max}}, \tag{3.24}$$

and similarly if the partner electron's initial subband is unknown (sum over j):

$$\Gamma_{i\to f}^{\text{e-e,max}} = \sum_j \Gamma_{ij\to f}^{\text{e-e,max}}. \tag{3.25}$$

The maximum total e-e scattering rate for an electron in subband i is then given by:

$$\Gamma_{i}^{\text{e-e,max}} = \sum_f \Gamma_{i\to f}^{\text{e-e,max}}. \tag{3.26}$$

Before we discuss the exact MC treatment of the e-e scattering, it is useful and instructive to examine how many parameters are actually unknown, and how we can most easily determine these parameters. We need to find the initial and final subbands and momenta of two electrons, i.e. four subband indices and 8 momentum components, or 12 parameters in total. We already know the initial subband and momentum of the followed electron, and the conservation of energy and momentum in the e-e scattering event yields another three parameters, so that means there are 6 unknowns left to determine with MC methods.

We start by determining the final subband of the followed electron in a similar way as described in Eq. 3.14. The next step is to randomly choose a partner electron (one initial subband and one initial momentum, i.e. 3 parameters) from among the other simulated electrons, which can easily be done by labeling the electrons 1,...,N and rolling the dice. The remaining unknown (final) subbands (2 parameters) are then easily found, again by applying Eq. 3.14. This leaves us with just one parameter to be determined, to be chosen from among the components of the final momenta (or derivatives thereof). Which one we chose is not very important, as the choice of one determines the others. In our implementation we picked the angle θ between $\mathbf{k}_f$ and the x-axis, which provides us with clearly defined range ($0 < \theta < 2\pi$) and a straightforward MC calculation.

By choosing the various subbands in the previous steps starting from $\Gamma_i^{\mathrm{e-e,max}}$, we have implicitly assumed a scattering rate of $\Gamma_{ij\to fg}^{\mathrm{e-e,max}}$. The goal of the next step is to determine the actual scattering rate for the proposed value of θ and compare it to $\Gamma_{ij\to fg}^{\mathrm{e-e,max}}$. Since the exchanged momentum q can easily be found from:

$$q(\theta)^2 = k_i^2 + k_f^2 - 2k_f k_i \cos\theta, \tag{3.27}$$

we can again apply the procedure described by Eq. 3.18 and find a suitable value for the final momenta of both electrons.

3.2 Algorithm

The Monte Carlo simulation consists of two major parts, namely the calculation of the various scattering rates and the main algorithm to determine the time evolution of the simulated electrons. There are some differences in implementation between the semi-classical simulation and the density matrix simulation, which are discussed in section 3.3. However, the general framework of both simulations is very similar. A basic flow chart of the algorithm is shown in Fig. 3-6.

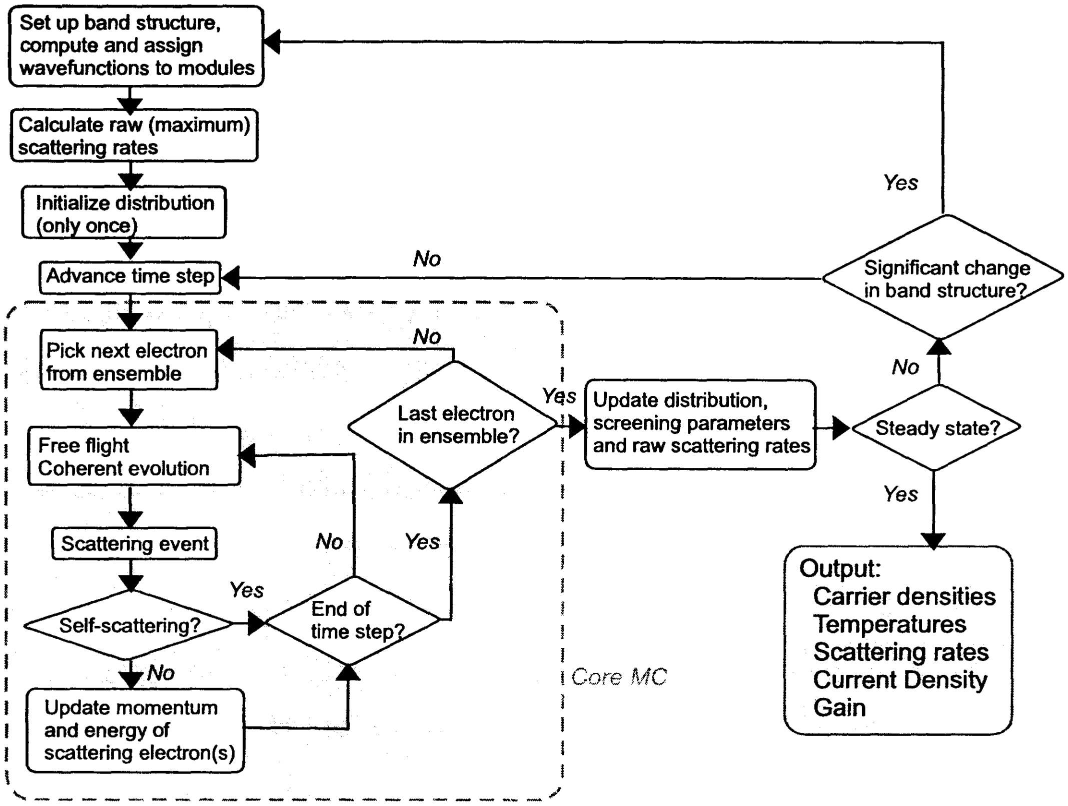

Figure 3-6: *Flow chart of the Monte Carlo simulation.*

In the simulation we keep track of the electrons belonging to (the subbands of) one module. In the THz QC structures of interest to us, the module length is generally comparable to or larger than the spatial extent of the wavefunctions, and as such the interaction between subbands is accurately described by same-module and next-neighbor scattering. By finding the wavefunctions belonging to a central module and its neighboring modules, we can calculate all relevant scattering parameters. Current continuity can then be enforced by reinjecting every electron that scatters out of the module into an equivalent level with an identical in-plane momentum **k**, reflecting the QC structure's periodic nature.

A proper choice of the basis wavefunctions is very important. Not only do these wavefunctions determine the various scattering rates, they also play a prominent role in the physical interpretation and intuition we can gain from the results. For the Monte Carlo simulation of QC structures, we have elected to work with two different choices of basis wavefunctions : the eigenstates of the Hamiltonian including all modules (for the semi-classical calculations) or the eigenstates of the Hamiltonian belonging to one single module (for the DM calculations), resulting in more localized wavefunctions. More details about this distinction can be found in the section discussing the DM simulation.

Once the basis wavefunctions are known, the matrix elements and maximum scattering rates for the electron-phonon, electron-impurity and electron-electron scattering are calculated as described earlier.

The number of simulated electrons N is chosen to allow for accurate calculation results while keeping the computation time as short as possible. For a semi-classical calculation, N_{sc}=10000 is usually sufficient to get satisfactory results, while for the density matrix simulations the initial number is N_{dm}=1000. However, as explained in the DM discussion, N_{dm} does not remain constant throughout the simulation, while N_{sc} does. In the semi-classical simulation, the simulation "particles" can be interpreted as single electrons, i.e. one particle equals one electron, while in the DM simulation the particles are in fact electron ensembles themselves which can be split into more, but smaller ensembles, giving rise to a change in the number of particles.

Each particle, numbered $p = 1, ..., N$, is characterized by its subband n_p and its momentum $\mathbf{k}_p$, which determine what scattering processes are available to that particle and what their rate is. To have a ready access to these different scattering rates, it is necessary to tabulate these rates as a function of n_p and $\mathbf{k}_p$. This is achieved most easily by defining a grid in momentum $\mathbf{k}$ space with step Δk in both the x- and y-directions:

$$\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} \;\rightarrow\; \mathbf{k}(i,j) = k_{x,i}\hat{\mathbf{x}} + k_{y,j}\hat{\mathbf{y}} \tag{3.28}$$

with

$$\begin{aligned} k_{x,i} &= i\Delta k, \;\; i = -N_k, N_k \\ k_{y,j} &= j\Delta k, \;\; j = -N_k, N_k \end{aligned} \tag{3.29}$$

where $2N_k$ is the number of grid nodes in the x- and y-directions. Usually N_k is taken to be 50 for a maximum kinetic energy E_k=250 meV, which translates into $\Delta k \approx 1.3 \times 10^7$ m^{-1}. We can then classify the different particles according to what area of $\mathbf{k}$ space they are in (and hence what scattering rates they are subject to). This allows us to use the previously calculated scattering rates and efficiently determine whether a scattering event is due to self-scattering or true scattering.

Finally, the electron distributions are initialized by randomly assigning the particles to one or more subbands with a thermalized distribution with temperature T_{latt}.

We are now ready to calculate the time evolution of the ensemble of particles. In this process, two time constants are of importance: the free flight time τ_f and the a time constant τ_{sys} that characterizes how quickly the electron distributions change. The free flight time determines how often we need to check whether a particle is involved in a (self-)scattering event; since the momentum of a particle is updated after every free flight, τ_f is also a good indicator of how recent a particle's momentum information is. In general, τ_f is a few tens of femtoseconds, mostly due to self-scattering. The "system change" time τ_{sys} on the other hand is most important for electron interactions that somehow depend on the other electrons in the system, e.g. electron-electron scattering and screening effects. These scattering rates which depend on the detailed electron distributions $f(k)$, need to be reevaluated every time the distributions change significantly. Since these distributions are, in turn, determined by the properties of the simulated particles, ideally the time evolution of all particles would be calculated simultaneously. In practice, $f(k)$ does not change too drastically over a time period of τ_{sys}, so we can time-evolve all particles in sequence, starting with the particle labeled "1", over a time step Δt not exceeding τ_{sys} (typically Δt=10-50 fs). During each time step, a particle can undergo multiple scattering events. Once the time evolution of last particle N is completed, the scattering rates and screening parameters are recalculated to account for the change in the electron distributions. The convergence of the simulation can then be checked by comparing the averaged electron distribution functions and current density over the previous time steps. After the $N_{\Delta t}$'th time step, we can define the average distribution function $\bar{f}_i^{\text{av}}(k)$ for subband i by averaging over 100 time steps:

$$\bar{f}_i^{\text{av}}(k) = \frac{1}{100} \sum_{i_{\Delta t}=N_{\Delta t}-99}^{N_{\Delta t}} f_i^{i_{\Delta t}}(k), \tag{3.30}$$

where $f_i^{i_{\Delta t}}(k)$ is the distribution function found after time step $i_{\Delta t}$. This average is compared to the average over the 100 previous time steps to determine convergence.

If there is no convergence, all particles are propagated over the duration of the next time step Δt, starting from where they left off, repeating the procedure explained before until the convergence requirements are met. During every time step, the number of scattered particles, their initial and final subbands and the responsible scattering mechanism are recorded. The current density is then found from the recorded flux of electrons scattering from the module into the next module, counting the electrons scattering back the previous module as negative.

The number of time steps necessary to reach a steady-state solution, depends sensitively on the details of the simulated structure and the initial distribution at simulation time t=0. In general, the required simulation time varies between 5 ps (for simple structures with high electron temperatures) to 50 ps (structures with poorly aligned subbands and low (initial) electron temperatures). For a time step

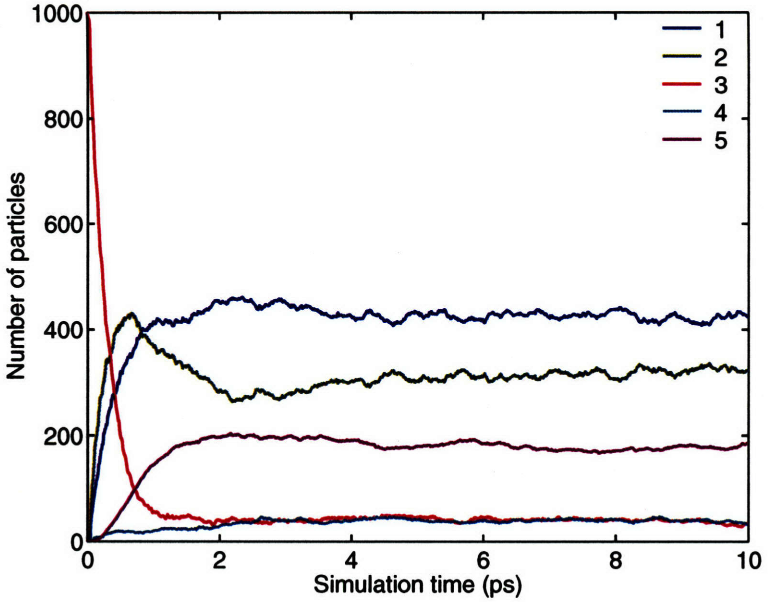

Figure 3-7: *Time evolution of the subband population densities in a DM simulation of FL177C-M5. The device is biased at injection anticrossing. All electrons are initially in subband n=3.*

Δt=10 fs this means that a regular MC simulation entails between 500 and 5000 iterations. An example of the evolution toward a steady-state solution is shown in Fig. 3-7 for a density-matrix simulation of the subband populations of FL177C-M5. Each particle represents a population density of 2.8×10^7 cm^{-2}. We can see that most subband populations reach steady-state after 2.5-5 ps. Note that it is necessary to continue the simulation for another few picoseconds (hundreds of timesteps) to obtain accurate results.

3.3 Numerical Implementation of DM simulation

The Monte Carlo method is very flexible and allows for a relatively simple and straightforward simulation of the equations of motion (2.76). On the other hand, the introduction of the density matrix formalism requires that we keep track of many more variables (e.g. the polarizations) than in the semi-classical case, and the computational requirements of a full-fledged implementation of these equations rapidly become very demanding. However, the proper choice of the basis wavefunctions allows us to develop an intuitive description of the transport in a QCL, which is partially coherent, and partially semi-classical.

In most of the demonstrated THz QCL structures, the levels within one module couple more strongly to each other than to levels in a different module, and a semi-classical description of transport within these regions is adequate. In order to reduce complexity, we restrict the use of coherent transport to model the transport between those modules, e.g. through the injector barrier, while retaining a semi-classical description for the transport inside each module. In this picture, current bottlenecks are described with the coherent density matrix model, while the transport through the rest of the device is described semi-classically. The basic implementation of this semi-classical part of the simulation was described in the previous sections and includes semi-classical electron-phonon (acoustic and LO), e-imp and e-e scattering.

The choice of the proper basis wavefunctions can be very important. This basis is used to calculate semi-classical scattering rates as well as to compose localized wave packets to model tunneling behavior and localization. As was explained in section 2.3.3, the use of spatially extended wavefunctions (as in semi-classical simulations) as basis wavefunctions fails to reproduce the experimentally observed tunneling behavior for transport through a barrier. To provide an intuitive picture of the electron states involved in transport, we choose a basis of wavefunctions which are localized within a module or submodule of the QCL under investigation. The thick injector barriers that confine a module form an obstruction for the electron transport, and we expect dephasing effects to be most prominent there. The choice of the basis wavefunctions as confined to either side of this barrier, makes it easy to describe and calculate resonant tunneling.

To find the localized wavefunctions ϕ_i^0, we consider a single, isolated module under bias, embedded in material with the same composition as the barriers. The wavefunctions in the previous (n=-1) and subsequent (n=1) modules are then found from:

$$\phi_i^n(x) = \phi_i^0(x - nl_{\text{mod}}) \tag{3.31}$$

with energy $E_i^n = E_i^0 + nqV_{\text{bias}}$, where $q = -1.6\times10^{-19}C$ is the electron charge, l_{mod} is the module length, and V_{bias} is the applied voltage per module. The interaction ΔV_{ij} between the localized wavefunctions ϕ_i^m and ϕ_j^n with respective in-plane momenta $\mathbf{k} = \mathbf{k_i}$ and $\mathbf{k_j}$ is then determined with a tight-binding model:

$$\Delta V_{ij}^n(\mathbf{k_i}, \mathbf{k_j}) = \int_{-\infty}^{+\infty} \phi_i^0(z)\Delta V(z)\phi_j^n(z)dz \ \delta_{\mathbf{k_i}\mathbf{k_j}}. \tag{3.32}$$

Here $\Delta V(z)$ is the difference in the confining potential between a single module and a superlattice composed of a repetition of this module:

$$\begin{aligned} \Delta V(z) = V_{\text{barr}} - \sum_{m<0} V_{\text{mod}}(z - ml_{\text{mod}}), &\qquad z < 0, \\ \Delta V(z) = 0, &\qquad 0 < z < l_{\text{mod}}, \\ \Delta V(z) = V_{\text{barr}} - \sum_{m>0} V_{\text{mod}}(z - ml_{\text{mod}}), &\qquad z > l_{\text{mod}}, \end{aligned} \tag{3.33}$$

where V_{barr} is the barrier potential and V_{mod} is the unbiased module potential profile. In practice, $\Delta V(z)$ is restricted to the influence of next-neighbor modules (n=-1,1) because QCL modules are usually very wide compared to the extinction length of the localized wavefunctions. Also note that $\Delta V_{ij}^1 = \Delta V_{ji}^{-1}$. The coherent intra-module transport due to ΔV_{ij}^0 is negligible compared to the scattering-mediated transport. It should be stressed again that in the absence of scattering, the transverse momentum $\mathbf{k}$ is conserved and $\Delta V_{ij}^n(\mathbf{k_i}, \mathbf{k_j})$ is non-zero only for states with an identical $\mathbf{k}$. In what follows we will abbreviate $\Delta V_{ij}^1(\mathbf{k_i}, \mathbf{k_j})\delta_{\mathbf{k_i}\mathbf{k_j}}$ with ΔV_{ij}. Note that this calculated interaction depends (weakly) on bias, as the changing potential and interactions within the module shift the wavefunctions around.

It should be pointed out that this approach yields a direct estimate of the anticrossing gap $\Delta_{ij} = 2|\Delta V_{ij}|$ between levels ϕ_i^0 and ϕ_j^1, and it makes this parameter easily accessible for investigation. In general, these calculated anticrossing gaps overestimate the value found from semi-classical calculations by 10-20%. Since the injector barrier thickness only affects the calculated Δ_{ij} and not the localized wavefunctions, this approach is very convenient to study the effect of barrier thickness on transport (see Section IV).

As mentioned before, we choose to adopt a "hybrid" strategy when including the density matrix formalism into the Monte Carlo simulation. Only the transport through the injector barriers is modeled in the DM formalism, the transport through the rest of the module is still handled semi-classically. This means that all scattering rates are calculated using Fermi's Golden Rule instead of a full density matrix description, and are simulated using a Monte Carlo approach. The transport through the barrier is handled by the quantum Liouville equation (Eqs. (2.76)-(2.78)), which includes the depopulation and pure dephasing scattering rates in the matrix $\mathcal{F}$. The solution to this equation exhibits oscillations on timescales varying from a few femtoseconds to tens of picoseconds, and a full Monte Carlo implementation of these equations would be very computationally intensive. Instead, we choose to analytically calculate the solutions to the equation of motion with a global phenomenological pure dephasing rate $\Gamma_{ij}^{\mathrm{pure}} = T_2^{-1}$ that applies to all subband states:

$$\frac{\partial \rho_{ij}}{\partial t} = \sum_{m,n} [\mathcal{L}_{ij,mn} - T_2^{-1} \delta_{im} \delta_{jn}] \rho_{mn}. \tag{3.34}$$

T_2 can be estimated from measurements of the spontaneous emission linewidth. By doing this, we describe the damping of the Rabi oscillation with two numerical methods: pure dephasing (T_2) is accounted for analytically and is assumed to be constant for the duration of the simulation; scattering dephasing (due to e-LO-phonon, e-e, and e-imp scattering) is calculated using a MC method and can (and will) vary with time. This scheme allows us to separate the time scales of pure and scattering dephasing, as the MC sampling rate is not affected by the pure dephasing time. The description of the time evolution of (an ensemble of) electrons between scattering events includes both coherent transport and pure dephasing. The practical implementation is straightforward. An arbitrary density matrix ρ can be described at time $t = t_0$:

$$\rho(t_0) = \sum_{k,l} c_{kl}(t_0) \rho_{kl}^0, \tag{3.35}$$

with

$$\rho_{kl}^0 = |\phi_k\rangle\langle\phi_l|. \tag{3.36}$$

Because the Liouville operator is a linear operator, we can describe the time dependence of a density matrix as the sum of the time evolutions of its components. If we write $\bar{\rho}_{kl}(t)$ for the solution to the equation of motion (including only pure dephasing scattering as discussed above, i.e. $\bar{\rho}_{kl}(t)$ is solved from Eq. (3.34) for the basis density matrices ρ^0_{kl}, we find:

$$\bar{\rho}_{kl}(t) = \sum_{m,n} c^{kl,0}_{mn}(t)\rho^0_{mn}, \tag{3.37}$$

and after a flight time Δt, ρ is transformed into:

$$\rho(t_0 + \Delta t) = \sum_{k,l,m,n} c_{kl}(t_0) c^{kl,0}_{mn}(\Delta t)\rho^0_{mn}. \tag{3.38}$$

The *a priori* unknown dephasing rates due to relaxation scattering can then be added during the MC simulation by setting the appropriate off-diagonal element ρ_{ij} to zero every second time a scattering event affecting levels i or j happens. This ensures that the dephasing rate due to relaxation scattering is half of $(\gamma_i + \gamma_j)$, as seen in Eq. (2.77). The affected diagonal element ρ_{ii} is adjusted accordingly as in the case of the semi-classical model.

A semi-classical Monte Carlo simulation deals with "integer" particles, i.e. every simulated particle represents an ensemble of electrons that is not broken up during the course of the simulation. In principle, all the particles in the ensemble can be name tagged and monitored during their transport process through the whole structure. The ensemble evolves and scatters as a whole, and at every point in time has a single well-defined momentum **k**. Whether a particle is viewed as a single electron or an ensemble of electrons is not important, since it does not affect the particle dynamics. Because the particles are indivisible, conservation laws dictate that their total number remains constant over the course of the simulation.

In the density matrix MC simulation, however, this is no longer the case, and we can no longer identify a particle with a single electron. Instead, each particle needs to be treated as an ensemble of electrons. At the beginning of the simulation, there is a limited number of particles with a specific energy and momentum, each represented by a density matrix. The absence of coherent interaction between two states with different **k**, makes it convenient to assign a different density matrix to every point in **k** space as their (coherent) time evolutions are independent. An ensemble ρ initially localized in one level i will quickly spread out across multiple levels due to coherent transport, while retaining its localization in **k** space:

$$\rho(0) = |\phi_i\rangle\langle\phi_i| \rightarrow \rho(t) = \sum_{k,l} c_{kl}(t)|\phi_k\rangle\langle\phi_l|. \tag{3.39}$$

However, relaxation scattering is calculated semi-classically and describes the transition of electrons from level i with initial momentum $\mathbf{k_i}$ (in the ensemble $\rho^{\mathbf{k_i}}$) to level j with final momentum $\mathbf{k_j}$. Consequently, a scattering event only affects the parts of $\rho^{\mathbf{k_i}}$ that refer to level i, namely the population density $\rho_{ii}^{\mathbf{k_i}}$ and the coherences $\rho_{ik}^{\mathbf{k_i}}(k \neq i)$, while the other elements remain the same:

$$\rho_{ii}^{\mathbf{k_i}} \rightarrow 0, \tag{3.40}$$

$$\rho_{ik}^{\mathbf{k_i}},\, \rho_{ki}^{\mathbf{k_i}}(k \neq i) \rightarrow 0\,(50\%\,\text{chance}). \tag{3.41}$$

The scattered electrons generally have a different momentum $\mathbf{k_j} \neq \mathbf{k_i}$, and need to be represented in a new density matrix $\rho^{\mathbf{k_j}}$, which initially consists of an electron wave packet localized in level j:

$$\rho^{\mathbf{k_j}} = |\phi_j\rangle\langle\phi_j|. \tag{3.42}$$

The electron population originally represented by one density matrix is now spread over two density matrices $\rho^{\mathbf{k_i}}$ and $\rho^{\mathbf{k_j}}$, and subsequent scattering will fragment them even further. The result is an ever increasing number of ensembles with different weights, spread out over **k** space. To counter this unbounded proliferation, we chose to group different ensembles according to their distribution in **k** space, by assigning them to "bins" chosen to represent a grid in **k** space. The different density matrices in one bin are very close in **k**, and can be approximately described with a single density matrix which is the weighted sum of all density matrices within this bin.

Chapter 4

Non-Equilibrium Green's Functions

4.1 Introduction : Non-Equilibrium Green's Functions

As discussed in the previous chapter, the density matrix formalism, as implemented, still leaves substantial room for improvement. The fine-tuning of the DM description would entail the introduction of many phenomenological dephasing times, which would in fact severely limit the model's envisioned use as a design tool. Instead, we chose to investigate the non-equilibrium Green's function (NEGF) formalism, which provides a more generalized microscopic theory for quantum transport. This approach is very comprehensive, and describes both quantum-coherent effects and scattering to an arbitrary order of perturbation. Furthermore, energy-resolved transport characteristics like level broadening are made readily available and are a natural part of the formalism, which is a marked improvement over the time-dependent density matrix formalism used in the previous chapter.

Non-equilibrium Green's Function theory [40, 41] is deeply rooted in the many-body theory, which describes how interactions between particles affect the behavior of a many-body system, such as nuclear systems, electrons in a lattice, plasmas and ferromagnetic media. The transport of electrons in QCLs, which is characterized by sub-picosecond scattering due to electron-phonon, electron-electron and electron-impurity scattering, is therefore a very suitable subject for the NEGF formalism. The particles in these systems exhibit very complicated behavior, which makes an accurate description difficult. Early attempts at solving these problems ranged from simply ignoring the many-body aspect (single particle approximations) to canonical transformations, in which a new basis is chosen to minimize the particle interactions. However, Feynman diagrams and path integrals, [42] methods developed originally for the Quantum Field Theory, together with the Green's function formalism provide a powerful way to attack the many-body problem.

As a real particle moves through interacting bodies, it causes a disturbance in its surroundings which travels with it. This cloud of agitated particles "dresses" the particle and can shield its interaction with other particles beyond the cloud. Good examples are the motion of an ion in an electrolytic fluid, or an electron in a metal or semiconductor. The central particle and its accompanying cloud of constantly changing particles behave as one entity and are known as a "quasi-particle". Many-body theory is a powerful instrument to translate the description of strongly interacting bare particles into a picture of weakly interacting quasi-particles, which is much easier to treat. Due to its composite nature and the interactions with its surroundings, the quasi-particle will in general be characterized by a lifetime τ and an energy E_{qp} and mass m_{qp} which are different from their counterparts E_{bare} and m_{bare} for the bare particle. The difference $E_{\text{qp}} - E_{\text{bare}} = E_{\text{self}}$ is called the "self-energy" of the quasi-particle. The particle interacts with the surrounding many-body system, creating a cloud, and the cloud in turn interacts with the particle, changing its energy. So, in a sense, the particle is interacting with itself.

The NEGF formalism treats these problems by examining what happens to an electron that is added to a system that is already in a non-equilibrium state, defined by a Hamiltonian $\hat{H}$ which may not be solved exactly:

$$\hat{H} = \hat{H}_0 + \hat{U} + \hat{H}_{\text{scatt}}, \tag{4.1}$$

In the Heisenberg representation, the unperturbed Hamiltonian can be written in term of the creation and annihilation operators $\hat{a}^\dagger_\alpha(t)$ and $\hat{a}_\alpha(t)$:

$$\hat{H}_0 = \sum_\alpha E_\alpha \hat{a}^\dagger_\alpha(t)\hat{a}_\alpha(t). \tag{4.2}$$

$\hat{H}$ has eigenstates $|\alpha\rangle$, and

$$\hat{U} = \sum_{\alpha,\beta} U_{\alpha\beta}(t)\hat{a}^\dagger_\alpha(t)\hat{a}_\beta(t) \tag{4.3}$$

describes nondiagonal parts of the Hamiltonian as well as the effect of electric fields, which generally leads to a coherent time evolution. $\hat{H}_{\text{scatt}}$ refers to elastic and inelastic interactions of electrons with phonons, ionic impurities, interface roughness potentials and other electrons. The language of second quantization allows for a convenient formal way to describe many scattering events, as the creation and destruction operators provide a good way to capture the dynamic interactions between the particles that form a quasi-particle's cloud. The various Green's functions are therefore defined as a function of these operators.

Our ultimate goal is to calculate various observables such as the occupation of state α, which can be expressed as the expectation value of the number operator:

$$f_\alpha(t) = \left\langle \hat{a}^\dagger_\alpha(t)\hat{a}_\alpha(t)\right\rangle. \tag{4.4}$$

Besides the one-particle density matrices $\left\langle \hat{a}^\dagger_\alpha(t)\hat{a}_\beta(t)\right\rangle$, higher order density matrices describe correlation effects and are, e.g., needed to accurately describe electron-electron scattering. [43]

In contrast to the density matrix formalism, where the temporal evolution of these observables is studied directly (all operators are taken at the same time t), Green's functions describe the correlation between two operators at times t and t'. As explained later, this extra degree of freedom is used to describe the energy spectrum and linewidth of a given state.

The exact derivation and justification of the various Green's functions and self energies is quite complicated and technical, and can be found in comprehensive reference texts such as written by Mahan. [44] Here I will summarize the key definitions and results that are important for this work, and I will try to provide a physical interpretation for each function as it is discussed.

4.2 Green's functions

In general, the Green's function formalism can give us the response at any point in space and time due to an excitation (addition of an electron or hole) of the system at any other point, including perturbations like scattering. This concept of Green's functions is very general and appears in many physical contexts in which the superposition priciple holds, such as circuit theory, electrostatics and electromagnetics. Since our focus is on quantum mechanics, the electronic system is described by the Hamiltonian H, and we define the Green's function G as: [45]

$$(E-H)G=S \tag{4.5}$$

where S is interpreted as an input (electronic excitation). If S is an impulse, we can formally write:

$$G=[E-H]^{-1}. \tag{4.6}$$

However, the inverse of a differential operator is not uniquely defined until we specify the boundary conditions. In general, two different Green's functions (retarded and advanced) are specified, each corresponding to a different boundary condition. This difference is best appreciated with a simple example of a one-dimensional system, excited at $x=x'$:

$$(E-H)G(x,x')=\delta(x-x'). \tag{4.7}$$

with a Hamiltonian given by:

$$H=-\frac{\hbar^2}{2m^*}\nabla^2 \tag{4.8}$$

If we look at Eq. 4.7, we see that it looks exactly like the Schrödinger equation, except for the source term $\delta(x-x')$. The most obvious solution is two waves originating at $x=x_0$ and traveling outward from the source:

$$G(x,x')=A\mathrm{e}^{\mathrm{i}k|x-x'|}. \tag{4.9}$$

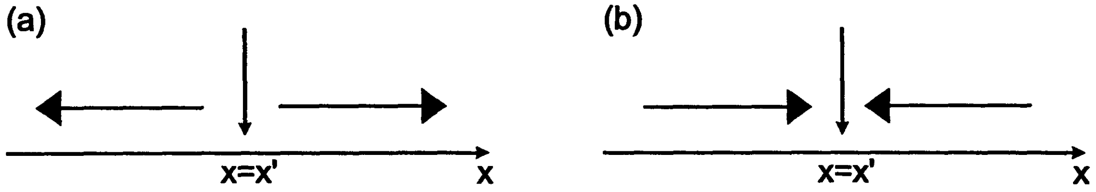

Figure 4-1: *Two possible solutions to Eq. 4.7. (a) The impulse at x=x' excites outgoing waves, propagating to infinity. (b) Waves incoming from infinity, meet at x=x' and are annihilated by the excitation.*

However, we can construct another solution as well:

$$G(x', x) = A\mathrm{e}^{-\mathrm{i}k|x-x'|}, \tag{4.10}$$

which corresponds to waves coming in from both sides and being annihilated at $x = x_0$ by the impulse. Note that, in this case, the waves travel in the opposite direction of the first solution. These two solutions, waves originating at the source and waves disappearing at the source (see Fig. 4-1) are commonly referred to as "retarded" and "advanced" Green's functions G^{ret} and G^{adv}, respectively. The terminology refers to the time domain, where $G^{\mathrm{ret}}(t, t')$ describes the time-retarded response, i.e. what happens at a time $t > t'$ due to an event at time t'.

The boundary conditions for the retarded Green's functions can be included explicitly into Eq. 4.7 by adding an infinitesimal imaginary part $\mathrm{i}\eta$ to the energy:

$$(E - H + \mathrm{i}\eta)\, G^{\mathrm{ret}}(x, x') = \delta(x - x'), \eta > 0, \tag{4.11}$$

The imaginary part makes the wavefunction grow indefinitely on the "wrong" side of the source, and hence ensures that the only acceptable solution consists of outbound waves. Similarly, adding $-\mathrm{i}\eta$ to the energy results in the advanced Green's function. We can then write formally:

$$G^{\mathrm{ret},0} = [(E + \mathrm{i}\eta) - H]^{-1}. \tag{4.12}$$

As stated before, this retarded Green's function describes (e.g., in the time domain) what the system looks like at time t after it was excited at time t', and no intervening excitations or interactions occurred. For this reason, the left-hand side of Eq. 4.12

is often referred to as the "free" or "undressed" retarded Green's function, which is denoted with the superscript "0". However, in the physical system we are dealing with, electron transport in QCLs, scattering and interaction with optical fields play an important role, and the influence of these interactions needs to be included in the expression for G^{ret}. Although the exact proof is beyond the scope of this introduction, the correct expression can easily be deduced by taking a closer look at another simple example.

Consider the problem of an electron propagating from point A to point B following a one-dimensional path, while undergoing scattering. The electron can get from A to B in many different ways, which involve a varying number of scattering events at a varying number of positions (not necessarily restricted to between A and B). As illustrated in Fig. 4.2, each of these paths can be broken up into consecutive periods of uninhibited travel (characterized by $G^{\text{ret},0}$) and a scattering event (with a probability P). So, the probability of finding the particle at point B is found by summing over all the different possible paths it can take: We can write this as: [42]

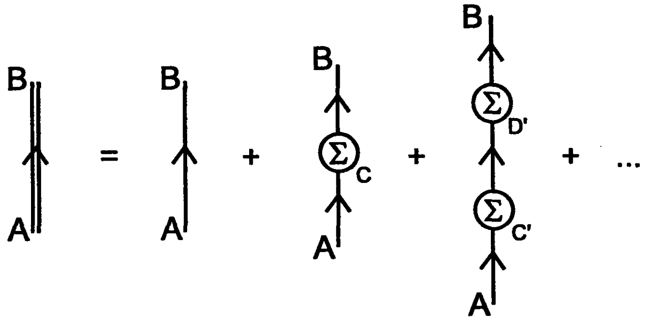

Figure 4-2: *Construction of the full propagator (double line) out of an infinite sum of partial propagators (single line). The dressed propagator includes all the effects of scattering.*

$$\begin{aligned} G^{\text{ret}}(\text{B,A}) &= G^{\text{ret},0}(\text{B,A}) + \sum_{C} G^{\text{ret},0}(\text{C,A})\, P(\text{C})\, G^{\text{ret},0}(\text{B,C}) \\ &\quad + \sum_{\text{C}',\text{D}'} G^{\text{ret},0}(\text{C}',\text{A})\, P(\text{C}')\, G^{\text{ret},0}(\text{D}',\text{C}')\, P(\text{D}')\, G^{\text{ret},0}(\text{B,D}') + \cdots \end{aligned} \tag{4.13}$$

Fourier transforming this expression simplifies the convolution terms on the right hand side, and we get:

$$\begin{aligned} G^{\text{ret}}(k_A,k_B) &= G^{\text{ret},0}(k_A,k_B)+\Sigma^{\text{ret}}[G^{\text{ret},0}(k_A,k_B)]^2+(\Sigma^{\text{ret}})^2[G^{\text{ret},0}(k_A,k_B)]^3+\cdots \\ &= \frac{G^{\text{ret},0}(k_A,k_B)}{1-\Sigma^{\text{ret}}G^{\text{ret},0}(k_A,k_B)}, \end{aligned} \tag{4.14}$$

where Σ^{ret} is the eigen-energy of the system and describes the effect of scattering on the electron's propagation. For a constant scattering rate $1/\tau$ the self-energy is given by $\Sigma^{\text{ret}} = -i\hbar/\tau$. Using the expression for $G^{\text{ret},0}$ from Eq. 4.12, we can write Eq. 4.14 as:

$$G^{\text{ret}} = \frac{1}{E-H-\Sigma^{\text{ret}}}, \tag{4.15}$$

where the infinitesimal term $i\eta$ is no longer needed in the denominator due to the presence of the imaginary self-energy Σ^{ret}. This expression is known as the Dyson Equation, and defines the relation between the electron propagator G^{ret}, which describes the electron dynamics, and the self-energy Σ^{ret}, which in turn describes how scattering affects an electron's coherent time evolution. It is straightforward to generalize this expression to describe a system of interacting states by replacing the scalar values with tensors:

$$\mathbf{G}^{\text{ret}} = [E\mathbf{I}-\mathbf{H}-\boldsymbol{\Sigma}^{\text{ret}}]^{-1}, \tag{4.16}$$

The Dyson equation is a generalization of Eq. 4.12 describing the "free propagator" $G^{\text{ret},0}$, and the poles of $G^{\text{ret}}(E)$ occur at (complex) energies which correspond to the energy difference between the excited states of the $(N+1)$-particle system (i.e., in our case, the biased QCL including the followed "test" electron) and the ground energy of the N-particle system. [42] As an example, we can look at a plane wave with energy E^k which has a retarded Green function given by:

$$G^{\text{ret}} = (E-E^k)^{-1}, \tag{4.17}$$

and the self-energy is zero due to the implied absence of scattering. We see that the pole of G^{ret} indeed corresponds with the energy of the electron. In our case, G^{ret}

contains easily accessible information about the energy and carrier lifetimes of the different subbands in a QCL. We can define the spectral function A which contains all the information about the energy level broadening: [45]

$$A_{\alpha_1\alpha_2}(\mathbf{k},E) = \mathrm{i}[G^{\mathrm{ret}}_{\alpha_1\alpha_2}(\mathbf{k},E) - G^{\mathrm{adv}}_{\alpha_1\alpha_2}(\mathbf{k},E)]. \tag{4.18}$$

This function plays the role of a generalized density of states.

In terms of creation and annihilation operators, the retarded and advanced Green's function can be written as: [46]

$$G^{\mathrm{ret}}_{\alpha_1,\alpha_2}(t_1,t_2) = -\mathrm{i}\Theta(t_1-t_2)\left\langle\left\{\hat{\mathrm{a}}_{\alpha_1}(t_1),\hat{\mathrm{a}}^{\dagger}_{\alpha_2}(t_2)\right\}\right\rangle, \tag{4.19}$$

$$G^{\mathrm{adv}}_{\alpha_1,\alpha_2}(t_1,t_2) = \mathrm{i}\Theta(t_2-t_1)\left\langle\left\{\hat{\mathrm{a}}_{\alpha_1}(t_1),\hat{\mathrm{a}}^{\dagger}_{\alpha_2}(t_2)\right\}\right\rangle = \left[G^{\mathrm{ret}}_{\alpha_2,\alpha_1}(t_2,t_1)\right]^{\dagger}, \tag{4.20}$$

where the brackets $\{\}$ denote the anti-commutator $\{\hat{\mathrm{a}},\hat{\mathrm{b}}\} = \hat{\mathrm{a}}\hat{\mathrm{b}} + \hat{\mathrm{b}}\hat{\mathrm{a}}$. In these definitions, the time-retarded and -advanced properties of these functions are expressed more directly through the step functions Θ. Note that G^{adv} and G^{ret} are Hermitian conjugates:

$$G^{\mathrm{adv}} = G^{\mathrm{ret}\dagger}. \tag{4.21}$$

Furthermore, it is straightforward to show that for a real (and hence symmetric) Hamiltonian H and a symmetric Σ^{ret}, G^{ret} and G^{adv} are symmetric as well. This will be the case in the simulations of the stationary transport, but not for the time-dependent gain calculations discussed in section 4.6. For a symmetric G^{ret}, the spectral density A can be related directly to G^{ret}:

$$A_{\alpha_1\alpha_2}(\mathbf{k},E) = -2\mathrm{Im}[G^{\mathrm{ret}}_{\alpha_1\alpha_2}(\mathbf{k},E)]. \tag{4.22}$$

4.3 Correlation function and $G^<$

As stated before, one of the main goals of the calculations is to obtain an estimate of the electron distribution function $f(\mathbf{k})$. In order to do this, we can generalize the expression for the time-varying density matrix $\rho(\mathbf{k},\mathbf{k}',t) = \left\langle\hat{\mathrm{a}}^{\dagger}_{\alpha}(\mathbf{k},t)\hat{\mathrm{a}}_{\alpha}(\mathbf{k}',t)\right\rangle$ to include correlations between the annihilation/creation operators at different times, and define the correlation function (or "lesser Green's function"): [45, 47]

$$G^{<}(\mathbf{k},\mathbf{k}',t,t') = \mathrm{i}\left\langle \hat{\mathrm{a}}^{\dagger}_{\mathbf{k}}(t')\hat{\mathrm{a}}_{\mathbf{k}'}(t)\right\rangle \tag{4.23}$$

Note that the density matrix can still be obtained from the correlation function by setting $t = t'$:

$$\rho(\mathbf{k},\mathbf{k}',t) = -\mathrm{i}\,[G^{<}(\mathbf{k},\mathbf{k}',t,t')]_{t=t'}, \tag{4.24}$$

In general, to describe time-varying transport we need to use the full two-time correlation function. However, since we are only interested in steady-state solutions, we can eliminate one of the time variables as the correlation function only depends on the time difference $(t - t')$ and use a Fourier transform to obtain:

$$G^{<}(\mathbf{k},\mathbf{k}',E) = \frac{1}{\hbar}\int dt\, \mathrm{e}^{\mathrm{i}Et/\hbar} G^{<}(\mathbf{k},\mathbf{k}',t+t',t') \tag{4.25}$$

The Fourier transform relationship between the energy E and the time difference $(t-t')$ can be most easily explained by considering a particle with energy E, evolving in time with a phase factor $\mathrm{e}^{-\mathrm{i}Et/\hbar}$. The correlation function is:

$$\psi(t)\psi^{*}(t') \propto \mathrm{e}^{-\mathrm{i}E(t-t')/\hbar}, \tag{4.26}$$

which suggests that the Fourier transform of the correlation function with respect to $(t - t')$ should yield the energy spectrum. In other words, the correlation function carries information regarding an energy level's linewidth and lineshape.

The (stationary) density matrix can then easily be found by integrating $G^{<}(\mathbf{k},\mathbf{k}',E)$ over all E (Eq. 4.25 with $t - t'$=0):

$$\rho(\mathbf{k},\mathbf{k}') = -\mathrm{i}\int \frac{dE}{2\pi} G^{<}(\mathbf{k},\mathbf{k}',E), \tag{4.27}$$

Because of the integration over E, the density matrix formalism retains no information about the electron energy spectrum, which makes it difficult to implement inelastic scattering processes (which change an electron's energy). In fact, our Monte Carlo implementation of the density matrix formalism implicitly assumes that the energy E^{elec} of an electron with momentum $\mathbf{k}$ in subband i is uniquely defined by $E^{\mathrm{elec}} = E_i^0+$

E^{kin} , where E_i^0 is its subband edge energy and $E^{\mathrm{kin}} = \hbar^2 k^2/2m^*$ is its semiclassical kinetic energy. This assumption is equivalent to treating the energy spectra as if they were delta functions centered on E^{elec}. It is of paramount importance to keep in mind that, in the NEGF formalism, there is no longer a one-to-one relationship between an energy state's momentum **k** and its kinetic energy, but that instead these two quantities should be treated separately. In the NEGF formalism, scattering and a finite lifetime lead to level broadening and an uncertainty in the energy of an electron; on the other hand in the semi-classical picture, scattering only leads to a change of a state's population while its energy remains well defined.

The diagonal elements of the correlation function (in subband index and momentum index) tell us the number of electrons occupying the corresponding energy and momentum. The semi-classical electron distribution function is then given by:

$$f(\mathbf{k}) = \rho(\mathbf{k},\mathbf{k}) = -\mathrm{i}\int \frac{dE}{2\pi} G^<(\mathbf{k},\mathbf{k},E). \tag{4.28}$$

In further analogy to the density matrix formalism, the off-diagonal elements of $G^<$ can be considered a generalized description of the degree of coherent interaction between different levels (polarization). For a subband α, the distribution function f can also be related to the spectral density A (generalized density of states):

$$\rho_{\alpha\alpha}(\mathbf{k},\mathbf{k}) = \frac{1}{2\pi\mathrm{i}}\int dE\, G^<_{\alpha\alpha}(\mathbf{k},E) = \int \frac{dE}{2\pi} A_{\alpha\alpha}(\mathbf{k},E) f(\mathbf{k},E). \tag{4.29}$$

Analogous to the electron correlation function, we can also define a "hole" correlation function $G^>$ which describes the hole distribution and phase correlations between holes. It is important to note that this function still refers to holes in the conduction band, and not in the valence band in the conventional sense, and as such defines the hole occupation number $1 - f(\mathbf{k})$:

$$G^>(\mathbf{k},\mathbf{k}',t,t') = -\mathrm{i}\left\langle \hat{\mathrm{a}}_{\mathbf{k}}(t)\hat{\mathrm{a}}^\dagger_{\mathbf{k}'}(t')\right\rangle \tag{4.30}$$

We can infer that, similar to the density matrix, $\mathrm{i}G^<$ and $-\mathrm{i}G^>$ are Hermitian.

As we will see in the discussion of the non-equilibrium Green's function formalism, during the calculations there is no need to explicitly derive $G^>$ at any point. The only correlation function necessary to determine important simulation results like the current density, subband population or gain is $G^<$.

4.4 Scattering and self-energy Σ

The self energy describes the interaction of an electron with itself and its environment, and can be considered the quantum-mechanical generalization of the semi-classical scattering matrix. The description of in- and outscattering gives rise to two different self-energies, $\Sigma^<$ and $\Sigma^>$ respectively, which are in turn closely related to their respective correlation functions (which is similar to stating that the in- and outscattering rates depend on the involved states' population densities). However, the exact form of that relationship is highly dependent on the described interaction.

For elastic, one-electron scattering, e.g. impurity or interface roughness scattering, the expression for the retarded ("ret") and lesser ("¡") self-energy follows a simple template in the Born approximation: [48, 46, 49]

$$\Sigma^{\mathrm{ret}/<,\mathrm{imp/rough}}_{\alpha_1\alpha_2}(\mathbf{k},E) = \sum_{\beta_1,\beta_2,\mathbf{k}'} \left\langle M^{\mathrm{imp/rough}}_{\alpha_1\beta_1}(\mathbf{k}-\mathbf{k}') M^{\mathrm{imp/rough}}_{\beta_2\alpha_2}(\mathbf{k}'-\mathbf{k}) \right\rangle \times G^{\mathrm{ret}/<}_{\beta_1\beta_2}(\mathbf{k}',E) \tag{4.31}$$

where $M^{\mathrm{imp}}_{\alpha_1\beta_1}(\mathbf{k}-\mathbf{k}')$ and $M^{\mathrm{rough}}_{\alpha_1\beta_1}(\mathbf{k}-\mathbf{k}')$ are matrix elements for impurity and roughness scattering, respectively (see section 2.2.3). The angle brackets denote an average over all possible distributions of impurities or variations in interface thickness. The expression for Σ^{adv} is identical to Eq. 4.31 with $G^{\mathrm{ret}} \rightarrow G^{\mathrm{adv}}$. Note that, for Σ^{ret}, this relation is very similar to Fermi's Golden Rule if we integrate the expression over E and we associate G^{ret} with a density of states. This tells us that we can interpret $\Sigma^{\mathrm{ret}}_{\alpha\alpha}(\mathbf{k},E)$ as the (intrinsic) scattering rate associated with the state $\mathbf{k}$ in subband α, and hence gives information about the state's lifetime. Similarly, $\Sigma^<$ is related to $G^<$ and is a measure for the electron flux associated with that state, i.e. the intrinsic rate weighted with the electron occupation. However, how do we determine whether this means scattering *into* or *out of* $\alpha, \mathbf{k}$? If we interpret $G^<$ as referring to the electron occupation number, we see that the right hand side of Eq. 4.31 is proportional to the number of electrons in states other than $\alpha, \mathbf{k}$, and hence $\Sigma^<$ describes the electron scattering flux *into* $\alpha, \mathbf{k}$. On the other hand, $\Sigma^>$ refers to the scattering of holes (as described by $G^{>}$) into $\alpha, \mathbf{k}$, or equivalently, the electron scattering flux *out of* $\alpha, \mathbf{k}$.

How do we relate these self-energies to the level lifetime Γ? As mentioned before, the retarded and advanced self-energies relate directly to the coherent propagation of an injected electron. It can easily be shown that: [45]

$$\Gamma = \mathrm{i}\left[\Sigma^{\mathrm{ret}} - \Sigma^{\mathrm{adv}}\right], \tag{4.32}$$

and as the retarded and advanced self-energies are Hermitian conjugates, the linewidth is given directly by the imaginary part of $2\Sigma^{\mathrm{ret}}$. Any time the electron is removed from the system, or interacts with its surroundings (the bath), the coherent evolution is over. This would seem to indicate that the appropriate lifetime is set by Σ^{out}. However, in case another electron tries to enter the system and is blocked by the electron that is already present, its coherent evolution is disturbed as well. This means that the rate at which electrons scatter into the state also contributes to its linewidth (cfr. the expression for dephasing scattering in the chapter about the density matrix formalism). Since the out-scattering rate of an electron is equivalent to the in-scattering rate of a hole, we can write:

$$\Gamma = \mathrm{i}\left[\Sigma^{<} + \Sigma^{>}\right] \tag{4.33}$$

These expressions for Γ are similar to the one for the spectral function A, and both describe, in somewhat different terms, the effects of scattering. It is therefore not surprising that A and Γ are linked by a simple transformation:

$$A = G^{\mathrm{ret}} \Gamma G^{\mathrm{adv}} = G^{\mathrm{adv}} \Gamma G^{\mathrm{ret}}. \tag{4.34}$$

For optical phonon scattering, the retarded self-energy can be written as a function of the phonon free propagators D^0, which describe the coherent evolution of non-interacting LO-phonons:

$$\begin{aligned}
\Sigma^{\mathrm{ret,LO}}_{\alpha_1\alpha_2}(\mathbf{k}, E) \;=\; & \mathrm{i}\sum_{\mathbf{k}'}\sum_{\beta_1,\beta_2} M_{\alpha_1\beta_1}(\mathbf{k}-\mathbf{k}') M_{\beta_2\alpha_1}(\mathbf{k}'-\mathbf{k}) \times \\
& \int \frac{dE_1}{2\pi}\Big[G^{\mathrm{ret}}_{\beta_1\beta_2}(\mathbf{k}', E-E_1) D^{\mathrm{ret}}(E_1) \\
& + G^{\mathrm{ret}}_{\beta_1\beta_2}(\mathbf{k}', E-E_1) D^{<}(E_1) + G^{<}_{\beta_1\beta_2}(\mathbf{k}', E-E_1) D^{\mathrm{ret}}(E_1)\Big]
\end{aligned} \tag{4.35}$$

with

$$D^{\text{ret}\,0}(E) = \frac{1}{E - \hbar\omega_{\text{LO}} + \text{i}0^+} - \frac{1}{E + \hbar\omega_{\text{LO}} + \text{i}0^+} \tag{4.36}$$

$$D^{\text{adv}\,0}(E) = \frac{1}{E - \hbar\omega_{\text{LO}} - \text{i}0^+} - \frac{1}{E + \hbar\omega_{\text{LO}} - \text{i}0^+} \tag{4.37}$$

$$D^{<0}(E) = -2\pi\text{i}\left\{n_{LO}\delta(E - \hbar\omega_{\text{LO}}) + (n_{LO}+1)\delta(E + \hbar\omega_{\text{LO}})\right\} \tag{4.38}$$

We then find for the full expression of $\Sigma^{\text{ret,LO}}$:

$$\begin{aligned}\Sigma^{\text{ret,LO}}_{\alpha_1,\alpha_2}(\mathbf{k},E) \;=\; & \sum_{\beta_1,\beta_2}\sum_{\mathbf{k}'} M_{\alpha_1\beta_1}(\mathbf{k}-\mathbf{k}')M_{\beta_2\alpha_1}(\mathbf{k}'-\mathbf{k})\Big[(n_{LO}+1)G^{\text{ret}}_{\beta_1\beta_2}(\mathbf{k}',E-\hbar\omega_{\text{LO}}) \\ & + n_{LO}G^{\text{ret}}_{\beta_1\beta_2}(\mathbf{k}',E+\hbar\omega_{\text{LO}}) + \frac{1}{2}G^{<}_{\beta_1\beta_2}(\mathbf{k}',E-\hbar\omega_{\text{LO}}) \\ & - \frac{1}{2}G^{<}_{\beta_1\beta_2}(\mathbf{k}',E+\hbar\omega_{\text{LO}}) + \text{i}\int\frac{dE_1}{2\pi}G^{<}_{\beta_1\beta_2}(\mathbf{k}',E-E_1) \\ & \times\left(\mathcal{P}\left\{\frac{1}{E-\hbar\omega_{\text{LO}}}\right\} - \mathcal{P}\left\{\frac{1}{E+\hbar\omega_{\text{LO}}}\right\}\right)\Big]\end{aligned} \tag{4.39}$$

where $n_{LO} = 1/[exp(E/k_BT) - 1]$, the equilibrium phonon occupation number at temperature T, introduces the only explicit temperature dependence in the whole Green's function formalism. Similarly for $\Sigma^{<,\text{LO}}$:

$$\begin{aligned}\Sigma^{<,\text{LO}}_{\alpha_1,\alpha_2}(\mathbf{k},E) \;=\; & \sum_{\beta_1,\beta_2}\sum_{\mathbf{k}'} M_{\alpha_1\beta_1}(\mathbf{k}-\mathbf{k}')M_{\beta_2\alpha_1}(\mathbf{k}'-\mathbf{k})\Big[n_{LO}G^{<}_{\beta_1\beta_2}(\mathbf{k}',E-\hbar\omega_{\text{LO}}) \\ & + (n_{LO}+1)G^{<}_{\beta_1\beta_2}(\mathbf{k}',E+\hbar\omega_{\text{LO}})\Big]\end{aligned} \tag{4.40}$$

4.5 Equations of motion

The temporal evolution of the Green's functions is given by an expansion of the simple Eq. 4.7: [50]

$$\begin{aligned}&\left(\text{i}\hbar\frac{\partial}{\partial t_1} - E_{\alpha_1}\right)G^{<}_{\alpha_1\alpha_2}(t_1,t_2) - \sum_{\beta}U_{\alpha_1\beta}(t_1)G^{<}_{\beta\alpha_2}(t_1,t_2) \\ &= \sum_{\beta}\int\frac{dt}{\hbar}\left[\Sigma^{ret}_{\alpha_1\beta}(t_1,t)G^{<}_{\beta\alpha_2}(t,t_2) + \Sigma^{<}_{\alpha_1\beta}(t_1,t)G^{\text{adv}}_{\beta\alpha_2}(t,t_2)\right],\end{aligned} \tag{4.41}$$

$$\left(\mathrm{i}\hbar\frac{\partial}{\partial t_2} - E_{\alpha_2}\right) G^{<}_{\alpha_1\alpha_2}(t_1,t_2) - \sum_{\beta} G^{<}_{\alpha_1\beta}(t_1,t_2)U_{\beta\alpha_2}(t_2)$$

$$= \sum_{\beta}\int\frac{dt}{\hbar}\left[G^{\mathrm{ret}}_{\alpha_1\beta}(t_1,t)\Sigma^{<}_{\beta\alpha_2}(t,t_2) + G^{<}_{\alpha_1\beta}(t_1,t)\Sigma^{\mathrm{adv}}_{\beta\alpha_2}(t,t_2)\right], \tag{4.42}$$

$$\left(\mathrm{i}\hbar\frac{\partial}{\partial t_1} - E_{\alpha_1}\right) G^{\mathrm{ret/adv}}_{\alpha_1\alpha_2}(t_1,t_2) - \sum_{\beta} U_{\alpha_1\beta}(t_1)G^{\mathrm{ret/adv}}_{\beta\alpha_2}(t_1,t_2)$$

$$= \hbar\delta(t_1-t_2)\delta_{\alpha1\alpha_2} + \sum_{\beta}\int\frac{dt}{\hbar}\Sigma^{\mathrm{ret/adv}}_{\alpha_1\beta}(t_1,t)G^{\mathrm{ret/adv}}_{\beta\alpha_2}(t,t_2), \tag{4.43}$$

$$\left(\mathrm{i}\hbar\frac{\partial}{\partial t_2} - E_{\alpha_2}\right) G^{\mathrm{ret/adv}}_{\alpha_1\alpha_2}(t_1,t_2) - \sum_{\beta} G^{\mathrm{ret/adv}}_{\alpha_1\beta}(t_1,t_2)U_{\beta\alpha_2}(t_2)$$

$$= \hbar\delta(t_1-t_2)\delta_{\alpha1\alpha_2} + \sum_{\beta}\int\frac{dt}{\hbar}G^{\mathrm{ret/adv}}_{\alpha_1\beta}(t_1,t)\Sigma^{\mathrm{ret/adv}}_{\beta\alpha_2}(t,t_2). \tag{4.44}$$

where $U(t)$ is generally used to describe the perturbations of the Hamiltonian that give rise to coherent motion, such as tight-binding matrix elements or the effects of electrical fields. The effects of scattering are included on the right hand side of the equations of motion, and are treated as source terms. Together with the expressions for the self-energy, these equations form a closed set of integro-differential equations which govern the temporal evolution of the Green's functions.

The relationship between G^{ret} and Σ^{ret} was already given in Eq. 4.15. The relation between $G^{<}$ and $\Sigma^{<}$ is referred to as the "Keldysh relation":

$$G^{<}_{\alpha_1\alpha_2}(t_1,t_2) = \sum_{\beta,\beta'}\int\frac{dt}{\hbar}\int\frac{dt'}{\hbar}G^{\mathrm{ret}}_{\alpha_1\beta}(t_1,t)\Sigma^{<}_{\beta\beta'}(t,t')G^{\mathrm{adv}}_{\beta'\alpha_2}(t',t_2) \tag{4.45}$$

which is obtained as a particular solution of Eqs. 4.41 and 4.42. As such, it holds true only in the limit of steady state.

Since we are interested only in stationary solutions, we can use Fourier transformations to simplify these equations:

$$G_{\alpha_1\alpha_2}(E) = \frac{1}{\hbar}\int dt\, \mathrm{e}^{\mathrm{i}Et/\hbar}G_{\alpha_1\alpha_2}(t+t_2,t_2), \tag{4.46}$$

$$G_{\alpha_1\alpha_2}(t_1,t_2) = \frac{1}{2\pi}\int dE\, \mathrm{e}^{-\mathrm{i}E(t_1-t_2)/\hbar}G_{\alpha_1\alpha_2}(E) \tag{4.47}$$

The equations of motion then become (in matrix notation):

$$(E\mathbf{I}-\mathbf{E})\cdot\mathbf{G}^{<}-\mathbf{U}\cdot\mathbf{G}^{<}=\mathbf{\Sigma}^{\mathrm{ret}}\cdot\mathbf{G}^{<}+\mathbf{\Sigma}^{<}\cdot\mathbf{G}^{\mathrm{adv}} \tag{4.48}$$

$$\mathbf{G}^{<}\cdot(E\mathbf{I}-\mathbf{E})-\mathbf{G}^{<}\cdot\mathbf{U}=\mathbf{G}^{\mathrm{ret}}\cdot\mathbf{\Sigma}^{<}+\mathbf{G}^{<}\cdot\mathbf{\Sigma}^{\mathrm{adv}} \tag{4.49}$$

$$(E\mathbf{I}-\mathbf{E})\cdot\mathbf{G}^{\mathrm{ret/adv}}-\mathbf{U}\cdot\mathbf{G}^{\mathrm{ret/adv}}=\mathbf{I}+\mathbf{\Sigma}^{\mathrm{ret/adv}}\cdot\mathbf{G}^{\mathrm{ret/adv}} \tag{4.50}$$

$$\mathbf{G}^{\mathrm{ret/adv}}\cdot(E\mathbf{I}-\mathbf{E})-\mathbf{G}^{\mathrm{ret/adv}}\cdot\mathbf{U}=\mathbf{I}+\mathbf{G}^{\mathrm{ret/adv}}\cdot\mathbf{\Sigma}^{\mathrm{ret/adv}} \tag{4.51}$$

and the Keldysh relation:

$$\mathbf{G}^{<}=\mathbf{G}^{\mathrm{ret}}\cdot\mathbf{\Sigma}^{<}\cdot\mathbf{G}^{\mathrm{adv}} \tag{4.52}$$

4.6 Gain

The information in $\mathbf{G}^{\mathrm{ret}}$ and $\mathbf{G}^{<}$ can also be used to find the gain and absorption spectra of the structure under bias. [51, 52] To do this, we consider the linear response of the stationary state described by the Green's functions to a small perturbation due to an optical field. In this context, we can consider the optical field to be too small to influence transport, and $\mathbf{G}^{\mathrm{ret}}$ and $\mathbf{G}^{<}$ do not depend on the applied optical field. This approach will yield the spontaneous emission spectrum, but cannot be used for the lasing regime.

The gain g and susceptibility χ are related by:

$$g(\omega)\approx-\frac{\omega}{c}\frac{\mathrm{Im}\left[\chi(\omega)\right]}{n_{\mathrm{B}}}, \tag{4.53}$$

where ω is the frequency of the optical field and $n_{\mathrm{B}}=\sqrt{\epsilon_{\mathrm{r}}}$ is the exciting mode's relative refractive index. The exciting optical field is polarized along the z-axis and propagates along y:

$$E(\mathbf{r},t)=\mathbf{e}_z\int\frac{d\omega}{2\pi}\mathcal{E}(\omega)e^{\mathrm{i}k(\omega)y-\mathrm{i}\omega t} \tag{4.54}$$

and results in a small time-dependent perturbation

$$\delta\hat{V}(\mathbf{r},t)=-\frac{e}{2m_e(z)}\left[\hat{\mathbf{p}}\cdot\mathbf{A}(\mathbf{r},t)+\mathbf{A}(\mathbf{r},t)\cdot\hat{\mathbf{p}}\right]+e\phi(\mathbf{r},t), \tag{4.55}$$

where $\hat{\mathbf{p}}$ is the momentum operator and $m_e(z)$ is the spatially dependent mass. The vector potential $\mathbf{A}$ and the scalar potential ϕ are related to the optical field through

$$\mathbf{A}(\mathbf{r},t) = \int \frac{d\omega}{2\pi}\frac{\mathcal{E}(\omega)}{\mathrm{i}\omega}e^{\mathrm{i}k(\omega)y-\mathrm{i}\omega t}\mathbf{e}_z, \tag{4.56}$$

$$\phi(\mathbf{r},t) = 0 \tag{4.57}$$

in the Coulomb gauge. In the long-wavelength approximation we neglect the spatial variation of the optical field throughout the structure and we can write in the energy representation:

$$\delta\hat{V}_{\alpha\beta}(\omega) = \mathrm{i}\frac{e\mathcal{E}(\omega)}{\omega}\left[\frac{\hat{p}_z}{m_e(z)}\right]_{\alpha\beta} = \frac{e\mathcal{E}(\omega)}{\hbar\omega}\left[\hat{H}_0,\hat{z}\right]_{\alpha\beta}. \tag{4.58}$$

The linear changes in the Green's functions and self-energies caused by $\delta V(t)$ represent the linear response of the nonequilibrium stationary state to the applied optical field. These changes can be written as:

$$\delta\mathbf{G}^{\mathrm{ret}}(\omega,E) = \mathbf{G}^{\mathrm{ret}}(E+\hbar\omega)\left[\delta\mathbf{V}(\omega)+\delta\mathbf{\Sigma}^{\mathrm{ret}}(\omega,E)\right]\mathbf{G}^{\mathrm{ret}}(E) \tag{4.59}$$

and

$$\delta\mathbf{G}^{\mathrm{adv}}(\omega,E) = \mathbf{G}^{\mathrm{adv}}(E+\hbar\omega)\left[\delta\mathbf{V}(\omega)+\delta\mathbf{\Sigma}^{\mathrm{adv}}(\omega,E)\right]\mathbf{G}^{\mathrm{adv}}(E), \tag{4.60}$$

and for the correlation function:

$$\begin{aligned}\delta\mathbf{G}^{<}(\omega,E) \;=\;& \mathbf{G}^{\mathrm{ret}}(E+\hbar\omega)\delta\mathbf{V}(\omega)\mathbf{G}^{<}(E)\\ &+\mathbf{G}^{<}(E+\hbar\omega)\delta\mathbf{V}(\omega)\mathbf{G}^{\mathrm{adv}}(E)\\ &+\mathbf{G}^{\mathrm{ret}}(E+\hbar\omega)\delta\mathbf{\Sigma}^{\mathrm{ret}}(\omega,E)\mathbf{G}^{<}(E)\\ &+\mathbf{G}^{\mathrm{ret}}(E+\hbar\omega)\delta\mathbf{\Sigma}^{<}(\omega,E)\mathbf{G}^{\mathrm{adv}}(E)\\ &+\mathbf{G}^{<}(E+\hbar\omega)\delta\mathbf{\Sigma}^{\mathrm{adv}}(\omega,E)\mathbf{G}^{\mathrm{adv}}(E).\end{aligned} \tag{4.61}$$

Note that $\delta\mathbf{G}^{\mathrm{adv}}(\omega,E) \neq \delta\mathbf{G}^{\mathrm{ret}\dagger}(\omega,E)$. The expression for $\delta\mathbf{\Sigma}(\omega,E)$ take the same functional form as in Eqs. 4.31 and 4.39-4.40, with $\mathbf{G}(E) \rightarrow \delta\mathbf{G}(\omega,E)$. For $\delta\mathbf{\Sigma}^{\mathrm{adv}}$, the expression for $\mathbf{\Sigma}^{\mathrm{ret}}$ is used with the replacement $\mathbf{G}^{<} \rightarrow -\delta\mathbf{G}^{<}$ and $\mathbf{G}^{\mathrm{ret}} \rightarrow -\delta\mathbf{G}^{\mathrm{adv}}$. The expressions for the self-energy must be evaluated self-consistently with those for

$\delta \mathrm{G}^{\mathrm{ret/adv/<}}$, and so another iterative loop, similar to the one for the nonequilibrium Green's functions, must be carried out.

From the change in the Green's functions we can find the change in the current density:

$$\delta J(\omega) = \delta J_0(\omega) + \delta J_{\mathrm{scatt}}(\omega) \tag{4.62}$$

where δJ_0 is the contribution from coherent current

$$\begin{aligned} \delta J_0(\omega) &= \frac{2e}{\hbar V} \sum_{\alpha\beta} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \left[\hat{H}_0, \hat{z}\right]_{\alpha\beta} \delta G^{<}_{\beta\alpha,\mathbf{k}}(\omega, E) \\ &\quad + \left[\delta \hat{V}(\omega), \hat{z}\right]_{\alpha\beta} G^{<}_{\beta\alpha,\mathbf{k}}(E) \end{aligned} \tag{4.63}$$

and $\delta J_{\mathrm{scatt}}$ can generally be neglected. The change in polarization is related to δJ through $\delta P = \mathrm{i}\, \delta J/\omega$, and hence the complex susceptibility can be found from:

$$\chi(\omega) = \frac{\delta P(\omega)}{\epsilon_0 \mathcal{E}(\omega)} = \frac{\mathrm{i}}{\epsilon_0} \frac{\delta J(\omega)}{\omega \mathcal{E}(\omega)} \tag{4.64}$$

From Eq. 4.53 we then find for the gain:

$$g(\omega) = -\frac{\mathrm{Re}[\delta J(\omega)]}{\epsilon_0 n_{\mathrm{B}} c \mathcal{E}(\omega)}. \tag{4.65}$$

4.7 Current

The Green's functions $G^<$ and G^{ret} allow us to extract a lot of information about the electron distribution, such as linewidth and state broadening, population density, density of states and lifetimes. However, an equally important parameter is the current density, which can easily be verified experimentally and therefore provides a good benchmark for the NEGF model. The current density can be calculated from the change over time of the position operator: [51]

$$J(t) = \frac{e}{V} \left\langle \frac{d\hat{z}}{dt} \right\rangle = \frac{e}{V} \frac{\mathrm{i}}{\hbar} \left\langle [\hat{H}, \hat{z}] \right\rangle. \tag{4.66}$$

If we split up the Hamiltonian $\hat{H} = \hat{H}_0 + \hat{H}_{\mathrm{scatt}}$, where $\hat{H}_0$ governs the coherent evolution (i.e. including off-diagonal elements that do not lead to a change in momentum

k) and $\hat{H}_{\rm scatt}$ describes elastic and inelastic scattering, the current density can be written as:

$$J(t) = \frac{e}{V}\frac{\rm i}{\hbar}\left\langle[\hat{H}_0{+}\hat{H}_{\rm scatt}]\right\rangle = J_0(t) + J_{\rm scatt}(t). \tag{4.67}$$

The first term on the right hand side describes the coherent current:

$$J_0(t) = \frac{e}{V}\frac{\rm i}{\hbar}\left\langle[\hat{H}_0{,}\hat{z}]\right\rangle = \frac{e}{\hbar V}{\rm Tr}\{[H_0, z]G^<(t,t)\} \tag{4.68}$$

In the energy representation this becomes:

$$J_0 = \frac{2e}{\hbar V}\sum_{\alpha,\beta}\sum_{\mathbf{k}}\int\frac{dE}{2\pi}[H_0, z]_{\alpha\beta}G^<_{\beta\alpha}(\mathbf{k}, E), \tag{4.69}$$

or using the density matrix ρ:

$$J_0 = \frac{2e}{\hbar V}\sum_{\alpha,\beta}\sum_{\mathbf{k}}[H_0, z]_{\alpha\beta}\rho_{\beta\alpha}(\mathbf{k}). \tag{4.70}$$

The expression for the scattering current is more involved:

$$\begin{aligned} J_{\rm scatt} \;=\; & \frac{2e}{\hbar V}\sum_{\alpha,\beta,\gamma,\delta}\sum_{\mathbf{k}}\int\frac{dE}{2\pi}[\Xi^<_{\alpha\delta}(\mathbf{k}, E) - z_{\alpha\gamma}\Sigma^<_{\gamma\delta}(\mathbf{k}, E)]G^{\rm adv}_{\delta\beta}(\mathbf{k}, E) \\ & +[\Xi^{\rm ret}_{\alpha\delta}(\mathbf{k}, E) - z_{\alpha\gamma}\Sigma^{\rm ret}_{\gamma\delta}(\mathbf{k}, E)]G^<_{\delta\beta}(\mathbf{k}, E), \end{aligned} \tag{4.71}$$

where the "transformed" self-energies Ξ are found by replacing $G_{\gamma\beta}$ by $\sum_\gamma z_{\alpha\gamma}G_{\gamma\beta}$ in the expressions for the self-energy. Note that although the coherent current contains no explicit reference to the scattering Hamiltonian, nevertheless its effects are included through $G^<$.

In order to get a more intuitive picture of the expression for $J_{\rm scatt}$, we consider only the diagonal elements of G and Σ in Eq. 4.71, and obtain:

$$\begin{aligned} J_{\rm scatt} \;=\; & \frac{2e}{\hbar V}\sum_{\alpha,\beta}\sum_{\mathbf{k}}\int\frac{dE}{2\pi} \\ & \times[\underbrace{G^<_{\beta\beta}\Sigma^{{\rm adv}(\alpha)}_{\beta\beta}}_{\rm Ia} + \underbrace{G^{\rm ret}_{\beta\beta}\Sigma^{<(\alpha)}_{\beta\beta}}_{\rm Ib}]z_{\beta\beta} + z_{\alpha\alpha}[\underbrace{\Sigma^{<(\beta)}_{\alpha\alpha}G^{\rm adv}_{\alpha\alpha}}_{\rm IIa} + \underbrace{\Sigma^{{\rm ret}(\beta)}_{\alpha\alpha}G^<_{\alpha\alpha}}_{\rm IIb}], \end{aligned} \tag{4.72}$$

where $\Sigma^{(\alpha)}$ refers to the contribution of $G_{\alpha\alpha}$ to the self-energy. The different terms in Eq. 4.72 can all be interpreted as scattering rates. The first term (Ia) consists of the product of the electron density in β ($G^<_{\beta\beta}$) with the raw scattering rate from β

into α ($\Sigma^{\text{adv},(\alpha)}_{\beta\beta}$), or in other words the electron flux $\Gamma^{\text{out}}_{\beta\to\alpha}$ out of β into α. Similarly, (IIb) describes the scattering rate $\Gamma^{\text{out}}_{\alpha\to\beta}$ out of α into β. Term (Ib) is the product of the density of states in β ($G^{\text{ret}}_{\beta\beta}$) and the in-scattering rate of electrons out of α into β ($\Sigma^{<(\alpha)}_{\beta\beta}$), i.e. the electron flux $\Gamma^{\text{in}}_{\alpha\to\beta}$ into β from α. (IIa) can be interpreted as $\Gamma^{\text{in}}_{\beta\to\alpha}$. Note that the out-scattering rates Γ^{out} are negative, while the Γ^{in} are positive. Since they describe the same current channel, the corresponding in- and out-scattering rates must be equal in magnitude and Eq. 4.72 can be rewritten as:

$$J_{\text{scatt}} \propto (\Gamma^{\text{out}}_{\beta\to\alpha} + \Gamma^{\text{in}}_{\alpha\to\beta})(z_{\beta\beta} - z_{\alpha\alpha}). \tag{4.73}$$

This expression tells us that the scattering current is proportional to the effective electron flux ($\Gamma^{\text{out}}_{\beta\to\alpha} + \Gamma^{\text{in}}_{\alpha\to\beta}$) multiplied with the distance between α and β, i.e. the net velocity and density of electrons moving from α to β.

4.8 Simulation

In order to decrease development time and to provide a solid frame of reference, the numerical implementation of the NEGF formalism was based in large part on the discussion of a similar simulation developed by Wacker *et al.* [46, 53, 52] Although in their work they neglect the momentum dependence of the scattering matrix elements, which leads to significant simplifications, the basic structure and algorithms of the actual calculations can easily be generalized to apply to our calculations. Furthermore, it is fairly straightforward to include the momentum-independent calculations in the simulation as a special case, adding the possibility to compare both approaches. The NEGF simulation includes the effects of electron-impurity, electron-interface roughness and electron-LO-phonon scattering in the Born approximation, and treats screening with a nonequilibrium multi-subband model identical to that used in the MC calculations. To address the absence of an adequate description of electron-electron (e-e) scattering and carrier thermalization, a tentative model for e-e scattering was developed and tested.

The ultimate goal of this simulation is to study the electron transport and optical gain properties in a quantum-cascade structure. A proper choice of the basis wave-

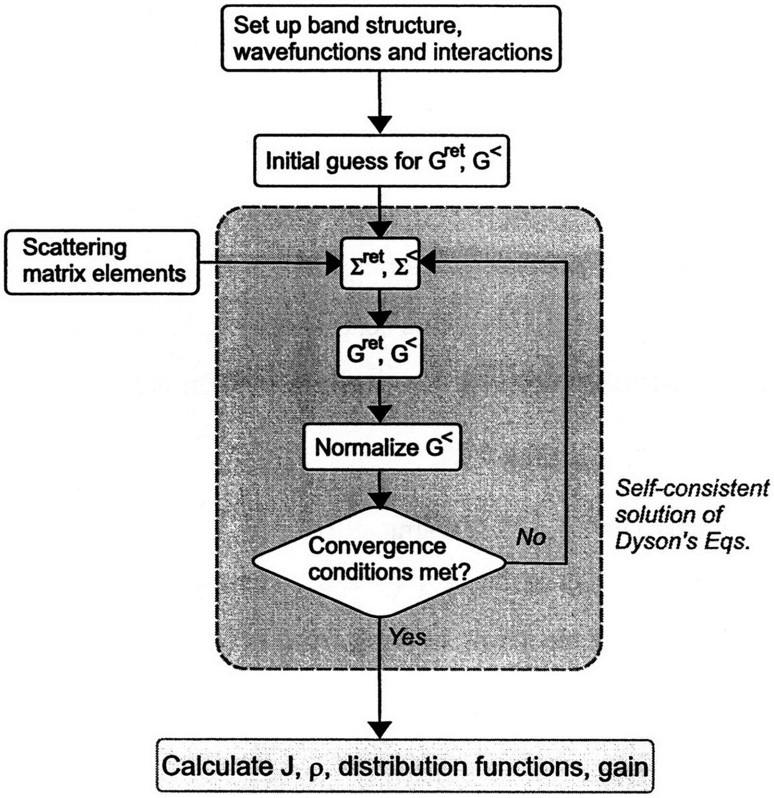

Figure 4-3: *Flowchart of the NEGF simulation.*

functions is of great importance, as it determines the scattering matrix elements and the physical picture we can get from the simulations. While the localized wavefunctions used for the density matrix Monte Carlo calculations can provide an intuitive physical picture, unfortunately the wavefunctions belonging to different modules are not orthogonal in this description, which renders many of the expressions for the current density and the gain invalid. In view of this, we chose to follow Wacker [46] and use Wannier functions as basis wavefunctions. While this choice of basis does not provide an intuitive picture, the resulting scattering matrix elements are independent of the applied bias, which allows for a simpler mathematical treatment of the calculation.

From simple initial guesses for G^{ret} and $G^{<}$, self-consistent solutions for the Green's functions and self-energies can then be obtained through an iterative calculation. These solutions can be used to find the current density, electron distribution function and gain spectrum of the investigated structure.

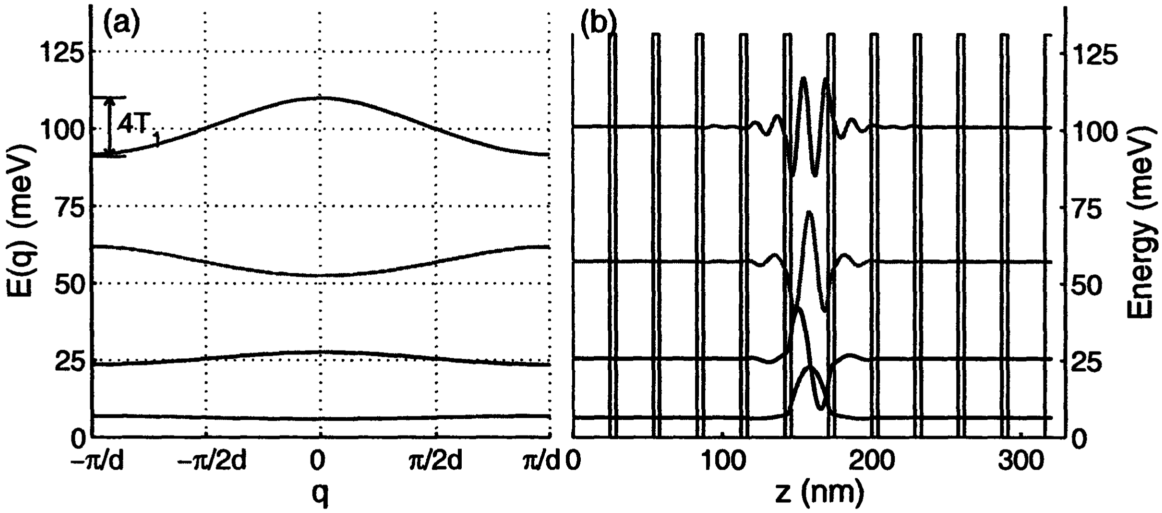

Figure 4-4: *(a) Dispersion relations for the 4 lowest-energy Bloch subbands of the superlattice. The miniband width is approximately given by* $4T_1$. *(b) Maximally localized Wannier wavefunctions, constructed from Bloch functions. The energy of each Wannier function corresponds to the middle of the Bloch miniband it was constructed from.*

The following sections provide more background and detail about the various aspects of the simulation, in particular the Wannier functions, the e-e scattering model and the practical implementation. As a practical example, we will examine the simulation results for a simple superlattice structure, consisting of a repetition of 25.1-nm-wide GaAs quantum wells separated by 4.0-nm-wide $\mathrm{Al_{0.15}Ga_{0.85}}$Asbarriers. Each barrier is doped to $n_{3\text{-D}} = 2.44 \times 10^{16}$ cm^{-3}, resulting in a 2-D electron density of 1×10^{10} cm^{-2}. The calculations were done for a lattice temperature T_{latt}=25 K.

4.8.1 Basis functions

Wannier functions

To construct the Wannier functions, [54, 55, 46] we start from the Bloch functions $\phi_q^\nu(z)$ of the superlattice potential, each characterized by its momentum q and grouped in minibands (index ν). Note that no external electric field is applied to this superlattice potential, and the obtained results will be independent of the bias voltage. The energy dispersion relation for a single miniband ν can be expanded in a Fourier series:

$$E^{\nu}(q) = E^{\nu} + \sum_{h=1}^{\infty} 2T_h^{\nu}\cos(hdq), \tag{4.74}$$

where the index h indicates the order of the Fourier term. In a typical case, the bands essentially have a cosine shape and only the term with T_1 is significant. Therefore, Eq. 4.74 is very similar to a standard nearest-neighbor coupling result with tight-binding strength T_1. The dispersion relation and the coupling parameters are illustrated in Fig. 4-4(a) and Table 4.1 for the superlattice. Note that the lower-energy subbands have a much smaller dispersion, corresponding to the larger effective confining potential they experience.

The Bloch wavefunctions are delocalized through the whole superlattice, which makes them a poor representation for the more localized wavefunctions that emerge when the superlattice is under a voltage bias. However, in analogy to the construction of localized wavepackets from a set of plane waves, it is possible to make a linear combination using the Bloch functions in each miniband ν to produce a localized Wannier function Ψ^{ν}:

$$\Psi^{\nu}(z - nd) = \sqrt{\frac{d}{2\pi}} \int_{-\pi/d}^{\pi/d} dq\, e^{-inqd} \phi_q^{\nu}(z), \tag{4.75}$$

where d is the module width and n indicates the position of the module with respect to a central module (n=0). These Wannier functions Ψ^{ν} each represent a different Bloch miniband ν, and are assigned the "central" energy E^{ν} of that miniband. It can be shown that these Ψ^{ν} are maximally localized when the phases of the Bloch functions are correlated in a carefully chosen manner, [54, 46] and for that choice Ψ^{ν}

Subband	E^{ν} (meV)	T_1 (meV)	T_2 (meV)
1	6.4005	-0.2369	0.0067
2	25.5272	0.9853	0.0310
3	57.1439	-2.3615	0.0623
4	100.8408	4.5564	0.1207

Table 4.1: *Energy parameters for the Bloch subbands of the superlattice. Since* $T_1 \ll T_2$*, the dispersion relations are nearly sinusoidal.*

will be real. The Wannier functions for the superlattice are illustrated in Figure 4-4(b). We can see that the Ψ^ν are largely confined to a single quantum well, with little of the wavefunctions leaking out to adjacent wells. However, unlike the tight-binding model, the wavefunctions are in fact infinitely extended and can be used to describe non-nearest-neighbor interactions with levels in distant wells.

Once the Ψ associated with one module are known, the functions for another module can easily be obtained by a proper translation. The Wannier functions thus found are orthonormal:

$$\int dz\, \Psi^\nu(z-nd)\Psi^\mu(z-md) = \delta_{\nu\mu}\delta_{nm}. \tag{4.76}$$

The Wannier functions belonging to the different modules (single wells in the case of our superlattice) of a quantum-cascade structure can then be used as orthonormal basis wavefunctions for the description of electron transport in that structure. As a first step, we need to construct the (unperturbed) Hamiltonian and describe the effect of an external bias field on the wavefunctions and energies.

Within second quantization, the creation and annihilation operators $a_n^{\nu\dagger}$ and a_n^ν associated with the Wannier function $\Psi^\nu(z-nd)$ can be found from the Bloch ladder operators in a way similar to Eq. 4.75. Taking into account only the next-neighbor terms with T_1, we can write for the Hamiltonian:

$$H_0 = \sum_{n,\nu}\left[E^\nu a_n^{\nu\dagger}a_n^\nu + T_1^\nu(a_{n+1}^{\nu\dagger}a_n^\nu + a_{n-1}^{\nu\dagger}a_n^\nu)\right]. \tag{4.77}$$

It is straightforward to show that an originally occupied Wannier state decays on a time scale of $\tau_{\text{Wannier}} = \hbar/2T_1$, which is very similar to the expression for the Rabi oscillation period found earlier. Therefore, T_1 fulfills a role very similar to the tight-binding elements in the density-matrix Monte Carlo simulation, with the exception that T only describes interactions between equivalent subbands in adjacent modules. For our superlattice example, we only take into account the bottom two subbands per module and find, for a central module and its adjacent modules (two levels each):

$$H_0 = \left[\begin{array}{cc|cc|cc} E^1 & & T_1^{(1)} & & & \\ & E^2 & & T_1^{(2)} & & \\ \hline T_1^{(1)} & & E^1 & & T_1^{(1)} & \\ & T_1^{(2)} & & E^2 & & T_1^{(2)} \\ \hline & & T_1^{(1)} & & E^1 & \\ & & & T_1^{(2)} & & E^2 \end{array}\right]. \tag{4.78}$$

Electric Field

The effects of an electric field F (along the growth direction z) can be included in the form of a perturbation $U = -eFz$. Note that the use of a non-orthonormal set of basis functions, as used in the density matrix MC simulation, would lead to complications when using the dipole matrix Z. The fact that wavefunctions belonging to different modules are not orthonormal is not an issue in the DM MC simulation because the only intermodule interaction is described through a tight-binding matrix element which does not rely on Z and is translation independent. However, in the NEGF formalism the off-diagonal elements of Z are used to describe the coupling between levels due to the electric field (and, as such, gain and coherent transport) and these are no longer translation independent for non-orthogonal wavefunctions:

$$Z^{\nu\mu}_{\alpha\beta} = \langle \Psi^{\nu}_{\alpha}(z)|z|\Psi^{\mu}_{\beta}(z)\rangle \neq \langle \Psi^{\nu}_{\alpha}(z)|(z+z_0)|\Psi^{\mu}_{\beta}(z)\rangle. \tag{4.79}$$

For this reason, the orthonormal Wannier functions are a better choice of basis functions for the used NEGF formalism than the localized wavefunctions described in the previous chapter.

For our superlattice, the dipole matrix can be written as:

$$Z = \left[\begin{array}{cc|cc|cc} Z_{11}^{00} - d & Z_{12}^{00} & Z_{11}^{0(-1)} & Z_{21}^{0(-1)} & & \\ Z_{12}^{00} & Z_{22}^{00} - d & Z_{12}^{0(-1)} & Z_{22}^{0(-1)} & & \\ \hline Z_{11}^{0(-1)} & Z_{12}^{0(-1)} & Z_{11}^{00} & Z_{12}^{00} & Z_{11}^{0(-1)} & Z_{21}^{0(-1)} \\ Z_{21}^{0(-1)} & Z_{22}^{0(-1)} & Z_{12}^{00} & Z_{22}^{00} & Z_{12}^{0(-1)} & Z_{22}^{0(-1)} \\ \hline & & Z_{11}^{0(-1)} & Z_{12}^{0(-1)} & Z_{11}^{00} + d & Z_{12}^{00} \\ & & Z_{21}^{0(-1)} & Z_{22}^{0(-1)} & Z_{12}^{00} & Z_{22}^{00} + d \end{array}\right], \tag{4.80}$$

where the diagonal elements are the expectation values for the position of the wavefunctions. The dipole elements between subbands in non-adjacent modules are usually negligible and are set to zero in our calculations. Note that Z is symmetric and that the off-diagonal elements are identical for any two adjacent modules. This periodicity and symmetry, also seen in H_0, is a reflection of the superlattice periodicity, and will also show up (in a slightly different form) in the Green's functions. These properties will allow us to reduce the amount of information we need to keep track of in the simulation.

Diagonalization of the Hamiltonian

While the choice of Wannier functions as basis wavefunctions substantially simplifies the simulations, it is not immediately helpful to gain an intuitive picture of the physics of the device under study. Furthermore, our choice of basis wavefunctions is also reflected in the scattering matrix elements and the Green's functions and self-energies, which renders them harder to interpret. This inconvenience can be lifted to a degree if we switch to a more intuitive basis, e.g. by diagonalizing the "coherent" Hamiltonian $H_0 - eFU$:

$$\mathbf{H'} = \mathbf{V} \cdot (\mathbf{H}_0 - eF\mathbf{U}) \cdot \mathbf{V}^{-1}, \tag{4.81}$$

where H' is the diagonalized Hamiltonian and V describes the linear transformation. The new basis functions Φ_i can then be found from the Wannier functions Ψ_n^ν:

$$\Phi_i = \sum_{\nu,n} V_{i,\nu n} \Psi_n^\nu \tag{4.82}$$

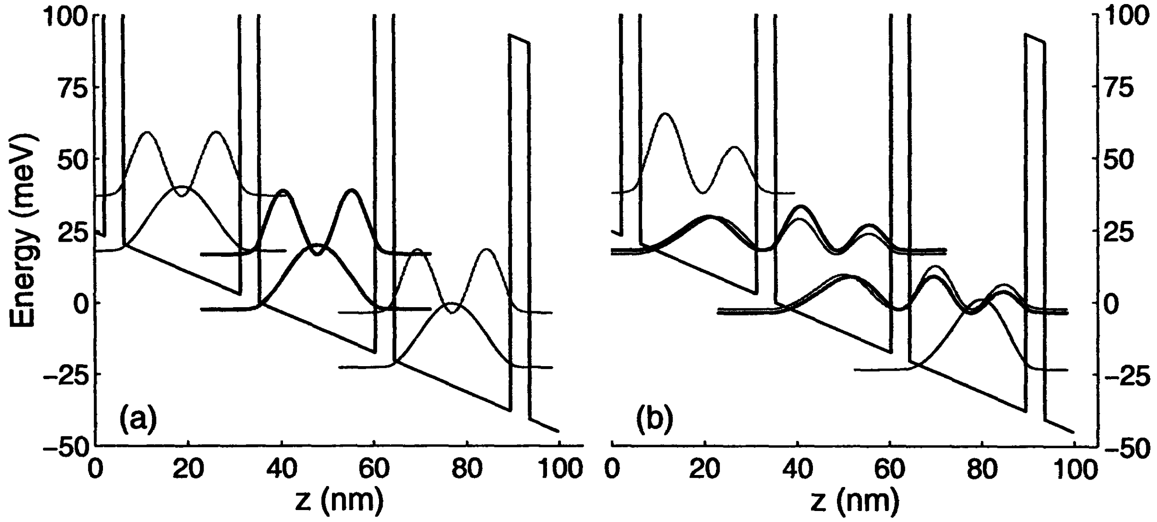

Figure 4-5: *(a) Band structure and square of the Wannier basis wavefunctions for a superlattice consisting of 25-nm-wide GaAs wells and 4.1-nm wide* $Al_{0.15}Ga_{0.85}As$*barriers. An external field of F=700 kV/m (18.9 mV/module) is applied. Every Wannier function is largely confined to a single well. The barriers are doped to* $n_{3-D} = 2.44 \times 10^{16}$ *cm*$^{-3}$*, resulting in a 2-D electron density of 1×10*^{10} *cm*$^{-2}$*. (b) Band structure and square of the transformed wavefunctions, for the same bias conditions as (a).*

The basis states obtained in this fashion are linear combinations of the Wannier functions, but have a very close resemblance to the spatially extended eigenstates used in the semi-classical Monte Carlo simulation. The limited number of Wannier basis states does not constitute a complete set, so there can be some differences between the extended eigenstates and the transformed wavefunctions. Figure 4-5 shows the Wannier wavefunctions and the transformed wavefunctions for the superlattice under resonant injection conditions (applied bias of 18.9 meV/module), for three adjacent wells. Note that the transformed wavefunctions are no longer confined to a single well but are spatially extended.

The Green's functions and self-energies can also be transformed to this more intuitive basis:

$$\mathbf{G} \to \mathbf{V} \cdot \mathbf{G} \cdot \mathbf{V}^{-1} \tag{4.83}$$

Similar transformations can also be applied to quantities derived from G^{ret} or $G^{<}$, such as the electron distribution function or the density matrix, which yields the transformed population densities.

4.8.2 Electron-Electron Scattering

To my knowledge, no working and practical model has hereto been proposed to handle electron-electron scattering in the NEGF formalism. In the absence of this scattering mechanism, there is no driving force to thermalize the electron distributions of the various subbands and electron scattering between closely spaced subbands could be severely underestimated. As such, the simulation results are affected both quantitatively and qualitatively, and may be less relevant when compared to experimental data. To remedy this, I devised a working model for e-e scattering and integrated it into the NEGF formalism. Although the used expressions do not have solid physical underpinnings, the e-e formalism is derived in an intuitive way and is shown to have the desired qualitative effect. I would like to thank Prof. Andreas Wacker for the many helpful discussions on this topic.

To get a working model for e-e scattering, we can start from the semi-classical expression for the scattering rate and infer its Green's function self-energy equivalent by making the proper adjustments. While this is not a scientifically well-founded derivation, the resulting equations and expressions are likely to reflect the more intuitive physics underlying the original semi-classical expression. If the dynamics introduced by the new terms in the Green's function formalism properly reproduce the behavior and effects we expect to see from electron-electron scattering, we can consider the additions successful, even if they are stopgap measures.

As mentioned earlier, to fit into the used NEGF formalism, e-e scattering needs to be described as a one-electron scattering mechanism, i.e. we can only keep track of the coherent evolution of one electron, whereas a proper Green's function description requires the time evolution of a system of two interacting electrons. This limitation implies that the phase information of one of the partner electrons will be discarded, i.e. in other words that electron can then be considered to be "semi-classical", and it has a well-defined energy $E = E^0 + E^k$ associated with its momentum. The other electron, the one of interest to us, can be described with the Green's function formalism. This picture essentially defines a background of semi-classical electrons for the followed "quantum-mechanical" electron to interact with. Another interesting possibility would be to model the e-e interaction after the electron-phonon interaction described in Eqs. 4.39 and 4.40, combining a description of an inelastic scattering process (from a one-electron point of view) with the coherent evolution of the partner particle (originally a phonon). While promising, this possibility was not explored further due to its mathematical complexity.

For the description of e-e scattering, expressions for both $\Sigma^{\mathrm{ret,e-e}}$ and $\Sigma^{<,\mathrm{e-e}}$ are required. The choice of an explicit form for one self-energy implies a correlated form for the other, as both self-energies merely describe different aspects of the same interaction. Therefore, we only need to construct an expression for either $\Sigma^{\mathrm{ret,e-e}}$ or $\Sigma^{<,\mathrm{e-e}}$, and the proper form of the other self-energy can then be inferred.

The most obvious self-energy to model is $\Sigma^{<,\mathrm{e-e}}$, as it has a straightforward physical interpretation that can easily be described in a semi-classical expression. $\Sigma^{<,\mathrm{e-e}}_{\alpha\alpha}(\mathbf{k}, E)$ tells us the electron flux into state $|\alpha, \mathbf{k}\rangle$ and energy E from all other states in the system, due to e-e interactions with any other electron in the system. As mentioned earlier, we are describing the scattering process of a "fully quantum-mechanical" electron with a "semi-classical" electron. Due to the e-e scattering event, the electrons exchange momentum $\mathbf{q}$ and can scatter from initial subbands (i, j) to final subbands (f, g):

$i, \mathbf{k}_i \longrightarrow f, \mathbf{k}_f$ (quantum-mechanical electron),

$j, \mathbf{k}_j \longrightarrow g, \mathbf{k}_g$ (semi-classical electron).

The initial and final energy of the semi-classical electron are uniquely defined as $E_j = E_j^0 + \hbar^2 k_j^2/2m^*$ and $E_g = E_g^0 + \hbar^2 k_g^2/2m^*$, respectively, while E_i and E_f do not have a direct relation to $\mathbf{k}_i$ and $\mathbf{k}_j$. E_j^0 and E_g^0 are the subband edge energies for subbands j and g, respectively.

There is a marked difference between the semi-classical and the quantum-mechanical description of e-e scattering, as illustrated in Fig. 4-6. In a semi-classical picture, a state's distribution function is described by a delta function in energy and momen-

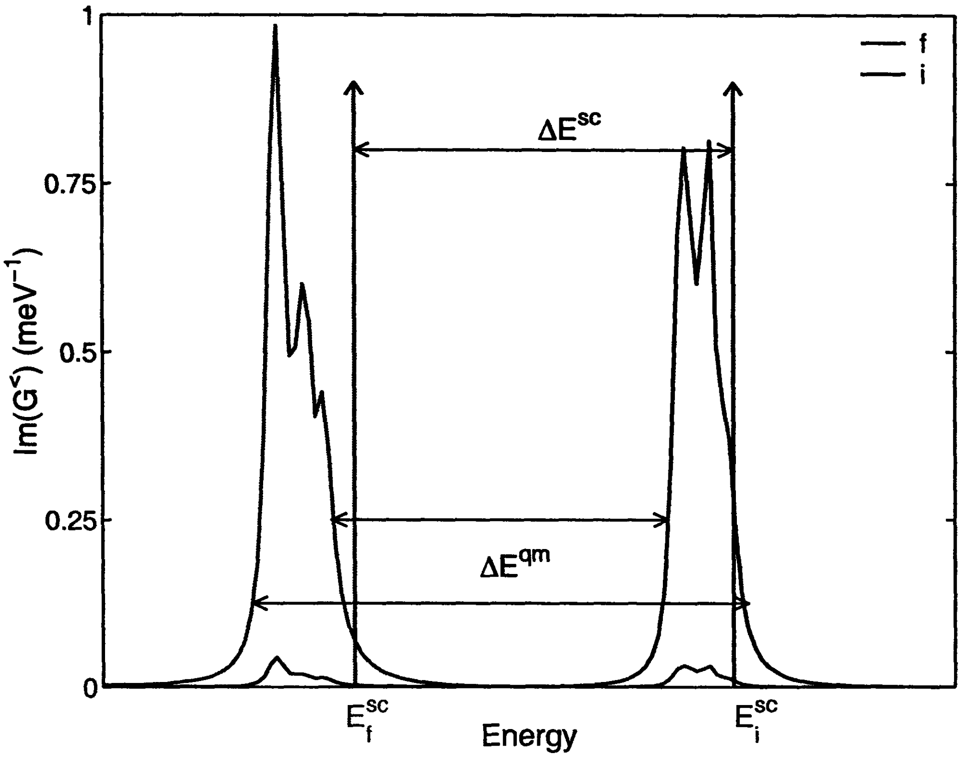

Figure 4-6: *Differences between the semi-classical and the full quantum-mechanical picture in allowed electron-electron scattering transitions between two states i and f. Shown are the semi-classical delta-function distributions and the quantum-mechanical equivalent. The only allowed semi-classical transition energy is ΔE^{sc}, while a wide range of energies ΔE^{qm} is possible for the quantum-mechanical transition.*

tum space, and scattering between two states i and f implies a fixed transition energy $\Delta E^{sc} = E_f - E_i$. Since the partner electron needs to have a transition available with the exact same energy and momentum transfer, the number of possible scattering partners is very limited. The quantum-mechanical distributions are still delta functions of momentum but smeared out in energy, and a large range of energies can be used to describe the transition. As can be seen in the figure, the decoupling of energy conservation and momentum conservation makes transitions possible which are not allowed in the semi-classical picture. If we look again at a transition between two states i and f, a larger range of partner electrons with an extended range of transition energies ΔE^{qm} can participate in e-e scattering. For our description of e-e scattering, where the transition energy and momentum are determined by a semi-classical partner electron, this means that we will have to sum over all these possible scattering partners, which adds to the numerical complexity of the calculation.

Let us now first examine the semi-classical scattering rate. In a fully semi-classical picture, the electron distributions are described by their distribution functions $f(\mathbf{k})$, and the electron flux into $|f, \mathbf{k}_f\rangle$ can be expressed using Fermi's Golden Rule:

$$W_f^{\mathrm{e-e}}(\mathbf{k_f}) = \frac{2\pi}{\hbar} \sum_{i,j,g} \sum_{\mathbf{k_i},\mathbf{k_j},\mathbf{k_g}} \int dE\, |V_{ijfg}(q)|^2 f_j(\mathbf{k_j})(1 - f_f(\mathbf{k_f}))(1 - f_g(\mathbf{k_g}))$$

$$\times \delta(E - E_i + E_j - E_f - E_g), \tag{4.84}$$

with the square of the e-e interaction potential:

$$|V_{ijfg}(q)|^2 = \frac{e^4}{(\epsilon_0 \epsilon_{\mathrm{GaAs}} A q)^2} |A_{ijfg}(q)|^2 \delta_{\mathbf{k}_i + \mathbf{k}_j, \mathbf{k}_f + \mathbf{k}_g}. \tag{4.85}$$

The form factor A_{ijfg} is given by:

$$A_{ijfg} = \int dz\, dz'\, \phi_i(z) \phi_f^*(z) \phi_j^*(z') \phi_g(z') e^{-q|z-z'|}, \tag{4.86}$$

and $q = |\mathbf{k_f} - \mathbf{k_i}|$.

To deduce $\Sigma^{<,\mathrm{e-e}}$ from Eq. 4.84, we retain the factors containing f_j and f_g (which describe the semi-classical electron) while replacing f_i and f_f with their Green's function counterparts, including the correlation function $G^<$ (Eq. 4.28). The energy delta function expresses the conservation of energy in the e-e scattering process and remains unchanged. Neglecting state-blocking, we can then write for $\Sigma^{<,\mathrm{e-e}}$:

$$\Sigma_{ff}^{<,\mathrm{e-e}}(\mathbf{k_f}, E) = \sum_{i,j,g} \sum_{\mathbf{k_i},\mathbf{k_j},\mathbf{k_g}} |V_{ijfg}(q)|^2 f_j(\mathbf{k_j}) G_{ii}^<(\mathbf{k_i}, E_i), \tag{4.87}$$

where $E = E_f$. The energy E_i can be determined by using conservation of energy:

$$E_i = E_f + E_g - E_j = E + E_g^0 - E_j^0 + \frac{\hbar^2}{2m^*}(k_g^2 - k_j^2), \tag{4.88}$$

and we can express k_g using momentum conservation:

$$k_g^2 = k_i^2 + k_j^2 + k_f^2 + 2k_i k_j \cos\theta_{\mathbf{k_i}\mathbf{k_j}} - 2k_i k_f \cos\theta_{\mathbf{k_f}\mathbf{k_i}} - 2k_j k_f \cos\theta_{\mathbf{k_f}\mathbf{k_j}}. \tag{4.89}$$

If we transform the sum over the momenta to an integral, and take into account the conservation of momentum in Eq. (4), we find:

$$\Sigma_{ff}^{<,\mathrm{e-e}}(\mathbf{k_f},E) = \frac{e^4}{(\epsilon_0\epsilon_{\mathrm{GaAs}}A)^2}\frac{A^2}{(2\pi)^4}\frac{m^{*2}}{\hbar^4}\sum_{i,j,g}\int dE_i^k\, d\theta_{\mathbf{k_f k_i}}\, dE_j^k\, d\theta_{\mathbf{k_f k_j}}$$

$$\times\left|\frac{A_{ijfg}(q(E_f^k,E_i^k,\theta_{\mathbf{k_f k_i}}))}{q(E_f^k,E_i^k,\theta_{\mathbf{k_f k_i}})}\right|^2 f_j(E_j^k)G_{ii}^{<}(E_i^k,E_i), \tag{4.90}$$

and $E_i(E_f^k, E, E_i^k, \theta_{\mathbf{k_f k_i}}, E_j^k, \theta_{\mathbf{k_f k_j}})$ is function of all the integration variables. Note that we only obtain an expression for the diagonal elements of the self-energy, as those are the only terms that have a direct semi-classical interpretation. However, Eq. 4.90 can easily be expanded to include the off-diagonal elements. Because of the added computational complexity and the fact that the diagonal self-energies seem to give adequate results, the off-diagonal elements were not calculated or examined any further.

To determine $\Sigma^{\mathrm{ret,e-e}}$, we take a closer look at the relation between the retarded and lesser self-energy for another non-elastic scattering process, namely LO-phonon scattering. In the expression 4.39 for $\Sigma^{\mathrm{ret,LO}}$, the terms containing $G^<$ and the principal-value integral are generally negligible compared to the terms with G^{ret}, and $\Sigma^{\mathrm{ret,LO}}$ can be approximated as a function of G^{ret} alone. Comparing the remaining terms in $\Sigma^{\mathrm{ret,LO}}$ with $\Sigma^{<,\mathrm{LO}}$, it can be seen that the term referring to LO-phonon absorption depends on $E + \hbar\omega_{\mathrm{LO}}$ for Σ^{ret} and $E - \hbar\omega_{\mathrm{LO}}$ for $\Sigma^<$, and vice versa for LO-phonon emission. For the rest, both expressions are identical. In other words, the only formal difference between the expressions, is that the exchanged energy $\hbar\omega_{\mathrm{LO}}$ flips its sign. This suggests that we can construct $\Sigma^{\mathrm{ret,e-e}}$ by copying $\Sigma^{<,\mathrm{e-e}}$, replacing $G^<$ with G^{ret}, and changing the sign of the exchanged energy $E_g - E_j$:

$$\Sigma_{ff}^{\mathrm{ret,e-e}}(\mathbf{k_f},E) = \frac{e^4}{(\epsilon_0\epsilon_{\mathrm{GaAs}}A)^2}\frac{A^2}{(2\pi)^4}\frac{m^{*2}}{\hbar^4}\sum_{i,j,g}\int dE_i^k\, d\theta_{\mathbf{k_f k_i}}\, dE_j^k\, d\theta_{\mathbf{k_f k_j}}$$

$$\times\left|\frac{A_{ijfg}(q(E_f^k,E_i^k,\theta_i))}{q(E_f^k,E_i^k,\theta_i)}\right|^2 f_j(E_j^k)G_{ii}^{\mathrm{ret}}(E_i^k,E_i'), \tag{4.91}$$

with

$$E_i' = E_f + E_g - E_j = E + E_g^0 - E_j^0 + \frac{\hbar^2}{2m^*}(k_g^2 - k_j^2). \tag{4.92}$$

As mentioned earlier, unfortunately the practical implementation of this expression leads to a very computation-intensive calculation, even for a simple structure. Let us consider the example of our 2-level superlattice, with only scattering between the two levels of one module. For an energy grid of 500 points (E), a momentum grid (E_k) of 100 points (see section 4.8.4)and 10 nodes in the grids for the different θ, the full calculation of the self-energy involves $2^3 \times 100^2 \times 10^2 = 8 \times 10^6$ summations to evaluate $\Sigma_{ff}^{<,e-e}(\mathbf{k_f}, E)$ for one value of f, $\mathbf{k_f}$ and E, or $2\times100\times500\times8\times10^6 = 8\times10^{11}$ operations to find all (diagonal) elements of $\Sigma^{<,e-e}$. This number will rise with the fourth power of the number of subbands involved in the calculation. To reduce the computational requirements of the e-e scattering expressions, several steps were taken to speed up the calculation by only retaining the most important terms in Eq. 4.91. For example, the very strong dependence of G on E and $f_j(E_j^k)$ on E_j^k allows us to restrict the integration to those values of E and E_j^k where these functions have their largest values. Additionally, e-e scattering is dominated by (bi-)intrasubband scattering and intersubband scattering of the type $(i, i) \rightarrow (f, f)$, so we can safely ignore other types and still retain the basic physical model of e-e scattering. These restrictions will alter the expected e-e calculation results, but should still express the influence of e-e scattering for the majority of the electrons (i.e. primarily at lower energy in a subband and at higher electron densities). For the superlattice calculations, only terms with $f_j(k_j) > 10^{-6}$ and $G(E^k, E) > \max(G)/100$ were taken into account. While more conservative cut-off values for f_j and G yielded slightly different results, the chosen values drastically reduced the calculation time while still incorporating the basic e-e scattering physics.

The effect of these extra self-energy terms becomes evident when we compare the (semi-classical) distribution function $f(k)$ for calculations with and without the e-e terms. Figure 4-7(a) shows $f(k)$ for the superlattice structure at anti-crossing, calculated without e-e scattering. The electron distributions are very non-thermal

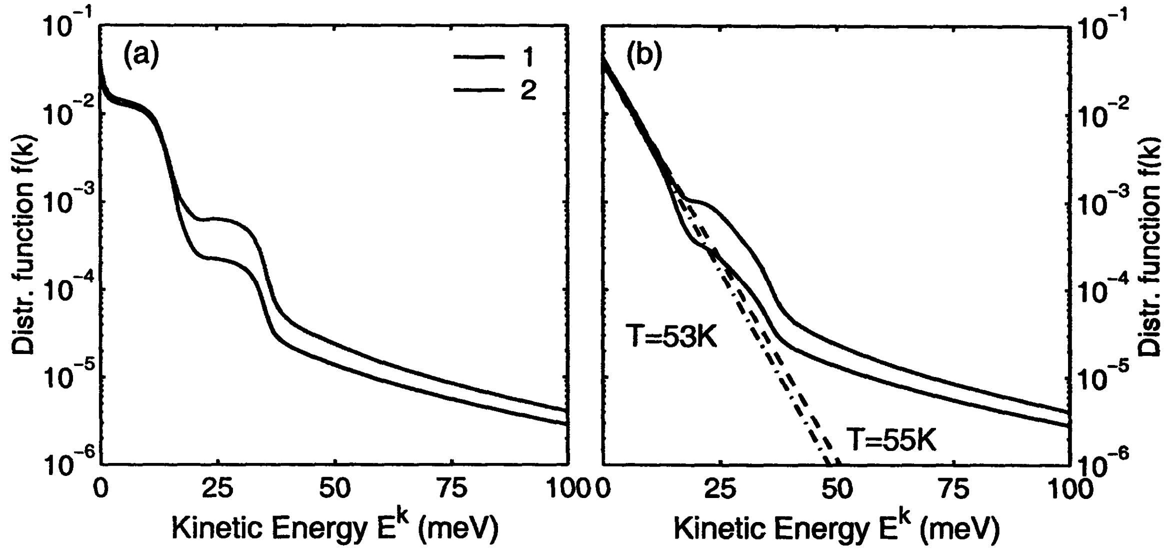

Figure 4-7: *(a) Distribution function $f(k)$ for both subbands of the superlattice for the NEGF simulation without e-e scattering. Note that both distributions are highly non-thermal. (b) Distribution function $f(k)$ for both subbands of the superlattice for the NEGF simulation including e-e scattering. The low-energy parts of the distributions are approximately thermalized.*

and exhibit large bumps around E^k=10 meV and E^k=30 meV. The origin of these features is not immediately clear. When the e-e terms are included (Fig. 4-7(b)) the low-energy range of the electron distribution is nearly thermalized, as indicated in the figure. Note that the corresponding temperatures for both subbands are very similar, which indicates a strong e-e interaction between the anticrossed states. For (kinetic) energies E^k >15 meV, corresponding to the onset of fast LO-phonon scattering from $n = 2$ to $n = 1$ ($E_{21} \approx 19$ meV), the distribution becomes non-thermal, although less so than in the simulation without e-e scattering. This indicates that the e-e scattering in those states is insufficiently fast to smoothen out the influence of LO-phonon scattering, i.e. $\tau_{\text{e-e}} \geq \tau_{\text{LO}} \approx 0.3$ ps. For the states with $E^k > \hbar\omega_{\text{LO}}$ this is all the more true, as intrasubband LO-phonon scattering becomes the dominant scattering mechanism. Note, however, that the thermal part of both electron distributions (up to $\sim$ 15 meV) comprises well over 90% of the total electron density. We can conclude that the main goal of the introduction of the e-e scattering self-energies, namely achieving a more thermalized electron distribution, has been attained.

4.8.3 Scattering Matrix Elements

As was the case for the Monte Carlo simulation, the scattering matrix elements for electron-phonon, electron-impurity, electron-interface roughness and electron-electron scattering play a central role in the calculations. In the used formalism, these matrix elements depend only on the Wannier functions, not on the voltage bias applied to the structure, which makes it possible to calculate the $M_{\alpha\beta}$ just once and use the results for a range of biases. The effects of an electric field are handled separately (U in Eqs. 4.41-4.44) and result in a change in the coherent current rather than the scattering current. This is in clear contrast with the Monte Carlo calculations, which start from bias-dependent wavefunctions and therefore end up with bias-dependent scattering matrix elements.

In general, the second stage of the simulation (the calculation of the Green's functions) is by far the most computation intensive and time consuming, and a proper choice of variables in the initial setup can greatly improve the whole algorithm's efficiency. Therefore it is usually advantageous to calculate and store the scattering matrix elements in a form that allows easy access and minimal processing during the iteration itself.

From Eqs. 4.31, 4.39 and 4.40, we see that the self-energies generally contain terms of the form:

$$\Sigma(\mathbf{k}, E) \sim \sum_{\mathbf{k}'} M_{\alpha_1\beta_1}(q) M_{\beta_2\alpha_2}(q) G_{\beta_1\beta_2}(\mathbf{k}', E), \tag{4.93}$$

where $q = |\mathbf{k} - \mathbf{k}'|$ is the exchanged momentum. Changing the summation into an integration we find:

$$\Sigma(\mathbf{k}, E) \sim \frac{m^*}{\hbar^2} \frac{A}{(2\pi)^2} \int dE^{k'}\, G_{\beta_1\beta_2}(\mathbf{k}, E) \int_0^{2\pi} d\theta M_{\alpha_1\beta_1}[q(\theta)] M_{\beta_2\alpha_2}[q(\theta)], \tag{4.94}$$

where $E^{k'} = \hbar^2 k'^2/2m$ is the kinetic energy associated with $\mathbf{k}'$ and θ is the angle between $\mathbf{k}$ and $\mathbf{k}'$. We can define a "interaction strength form factor", encompassing the integration over θ:

$$B_{\alpha_1\beta_1\beta_2\alpha_2}(E^k, E^{k'}) = \int_0^{2\pi} d\theta M_{\alpha_1\beta_1}[q(\theta)] M_{\beta_2\alpha_2}[q(\theta)]. \tag{4.95}$$

This form factor effectively describes the scattering strength (for the studied interaction) between states $|\alpha_1, \mathbf{k}\rangle$ and $|\alpha_2, \mathbf{k}'\rangle$. It is easy to show that $B_{\alpha_1\beta_1\beta_2\alpha_2}(E^k, E^{k'}) = B_{\alpha_1\beta_1\beta_2\alpha_2}(E^{k'}, E^k)$ and $B_{\alpha_1\beta_1\beta_2\alpha_2}(E^k, E^{k'}) = B_{\alpha_2\beta_2\beta_1\alpha_1}(E^{k'}, E^k)$. Note that B does not depend on the Green's functions, and can be evaluated without any prior knowledge of G. It is also worth remarking that this interaction strength is not limited by any cut-off energy or other restriction due to conservation of energy. The energies E^k and $E^{k'}$ are no more than a convenient way to refer to the momenta k and k', and the real energy dependence is contained in the parameter E and the Green's functions. The energy spectra of $G^<$ and G^{ret} serve to determine which matrix elements contribute to a certain transition, much like the energy-conserving delta function did in the semi-classical description.

The self-energy can then be expressed as:

$$\Sigma(E^k, E) \sim \frac{m^*}{\hbar^2} \frac{A}{(2\pi)^2} \int dE^{k'}\, G_{\beta_1\beta_2}(\mathbf{k}', E) B_{\alpha_1\beta_1\beta_2\alpha_2}(E^k, E^{k'}). \tag{4.96}$$

Using the form factor leaves us with only one integration in the expression for the self-energy, which is a vast improvement over Eq. 4.94 as the angle integration does not need to be repeated every time Σ is calculated. However, this improvement comes at a cost: instead of having to keep track of just $M_{\alpha\beta}(q)$ (3 dimensions) in Eq. 4.94, now we need 6 dimensions for $B_{\alpha_1\beta_1\beta_2\alpha_2}(E^k, E^{k'})$. While the resulting use of storage memory may not be optimal, the gain in calculation efficiency and speed far outweighs the disadvantages.

Calculation of B

As mentioned earlier, the calculation of B can be a time and memory consuming process, and it can be very advantageous to explore how we can make this process more efficient. To do this, we take a closer look at the exchanged momentum q (Fig. 4-8):

$$q = \sqrt{k^2 + k'^2 - 2kk'\cos(\theta)}. \tag{4.97}$$

Because the transport perpendicular to the growth direction is isotropic, we can choose to point $\mathbf{k}$ along the k_x-axis without losing the generality of the end results.

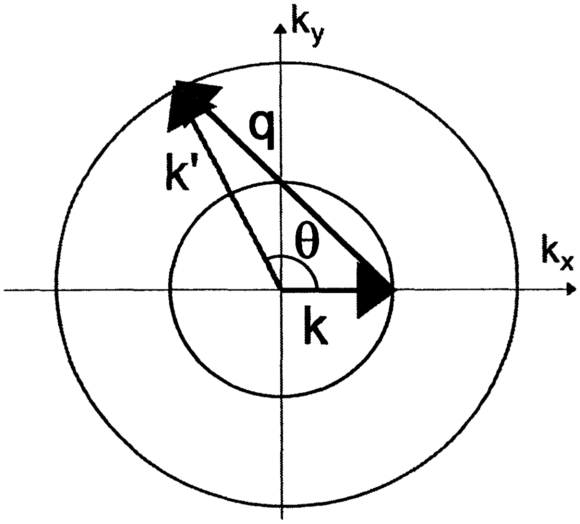

Figure 4-8: *Graphical representation of the relation between* **k**, **k**′ *and* **q**. *The circles represent the possible end points of the momentum vectors belonging to states with (kinetic) energies* E^k *and* $E^{k'}$.

We see that varying θ makes $\mathbf{k}'$ describe a circle while $\mathbf{k}$ remains fixed. From Fig. 4-8 and Equation 4.97, it is clear that q will change most rapidly as function of θ when $\cos(\theta) \approx 1$, i.e. for small values of $|\theta|$ and when q is minimal. Furthermore, the matrix elements also tend to exhibit a stronger q-dependence for small values of q, as shown in Figures 4-9(a) and 4-10(a) for several key transitions in our superlattice. Especially the elements involving intersubband transitions ("1122" in the figures) change very rapidly for small values of q. It follows that M can be very sensitive to θ as well, and this needs to be taken into account when evaluating B numerically:

$$B_{\alpha_1\beta_1\beta_2\alpha_2}(E^k, E^{k'}) \approx 2\sum_i \Delta\theta M_{\alpha_1\beta_1}[q(\theta_i)]M_{\beta_2\alpha_2}[q(\theta_i)], \quad \theta_i = (i-1)\Delta\theta,\ i = 1, N \tag{4.98}$$

where $N = \pi/\Delta\theta$ is the number of steps in the θ-grid. However, since we are using uniform steps in θ to step through the summation, the step size $\Delta\theta$ needs to be chosen small for Eq. 4.98 be accurate for small θ (and q), while a much larger $\Delta\theta$ would be sufficient for larger θ. Therefore, choosing a non-uniform grid with increasing $\Delta\theta$ can be much more efficient. A good choice is a linearly increasing grid step:

$$\Delta\theta_i = i\Delta\theta_0 \rightarrow \theta = \frac{i(i+1)}{2}\frac{2\pi}{N(N+1)}, \tag{4.99}$$

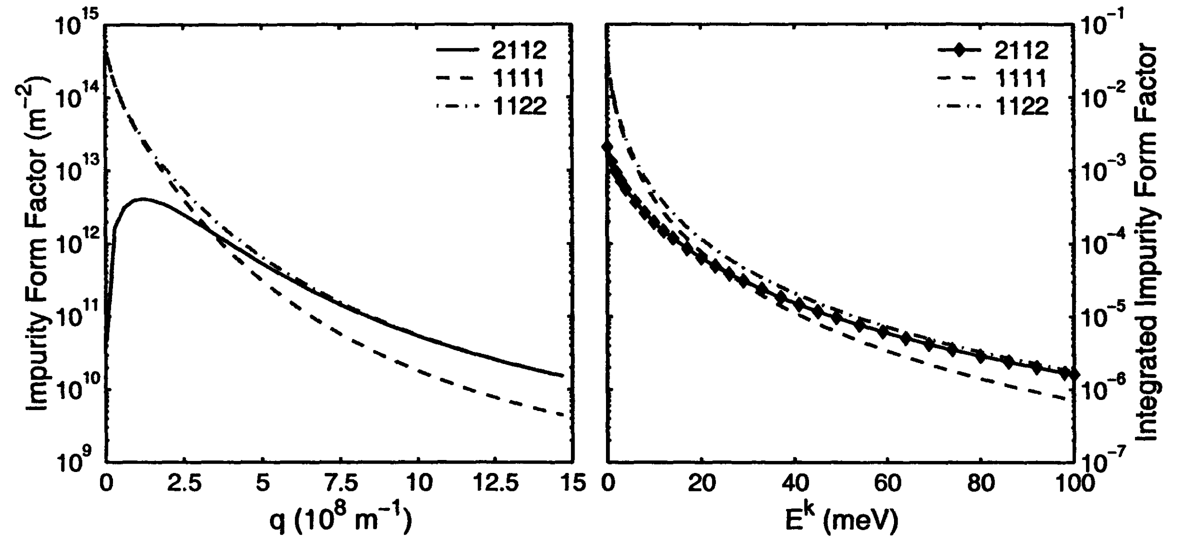

Figure 4-9: *(a) Impurity scattering form factor as a function of q for different transitions. (b) Integrated impurity scattering form factor in function of* E^k *for* $E^{k'}=0$ *meV.*

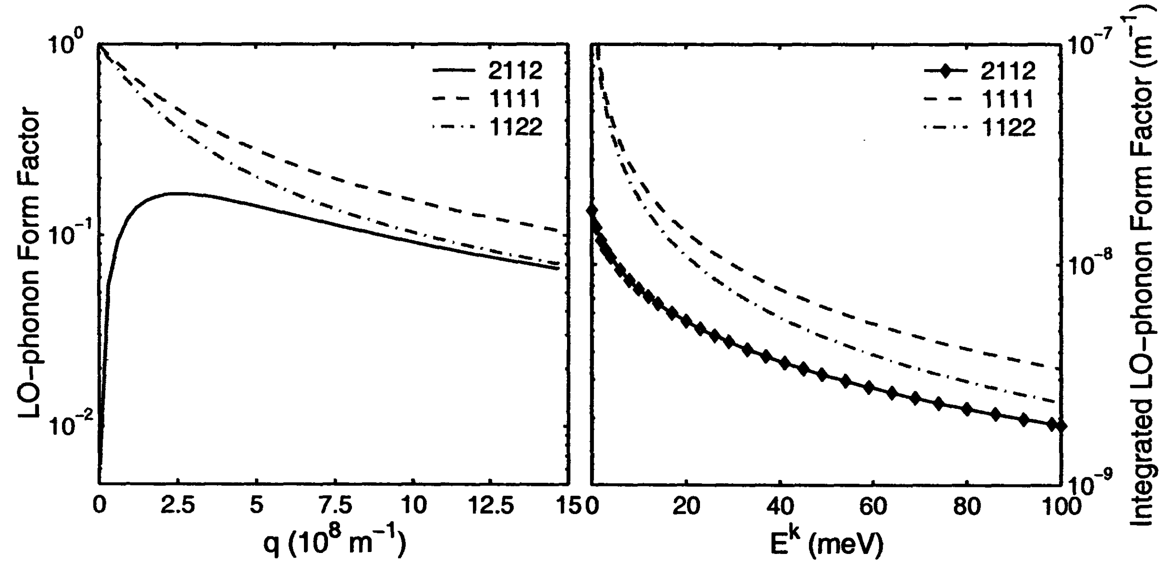

Figure 4-10: *(a) LO-phonon scattering form factor as a function of q for different transitions. (b) Integrated LO-phonon scattering form factor in function of* E^k

where the minimum grid step $\Delta\theta_0 = 2\pi/N(N+1)$ is much smaller than the step $\Delta\theta = \pi/N$ for a uniform grid with the same N. While the maximum step size $2\pi/(N+1)$ is almost twice as large as in the uniform grid, it occurs at relatively large q values where the matrix elements are less sensitive to q and the step size is less critical. To match the largest step size with $\Delta\theta$, it is sufficient to roughly double the number of nodes in the non-uniform grid; however, to match the smallest step in the non-uniform grid, we need to square the number of nodes in the uniform grid. This illustrates that the non-uniform grid is the better choice.

4.8.4 Grid

The various Green's functions $G_{\alpha\beta}(\mathbf{k}, E)$ and self-energies $\Sigma_{\alpha\beta}(\mathbf{k}, E)$ are functions of five variables, i.e. two subband indices, the two components of the momentum $\mathbf{k}$ and the energy E. Fortunately, we can assume that G and Σ are isotropic perpendicular to the growth direction, so $\mathbf{k}$ can be replaced with the momentum magnitude k, or equivalently the kinetic energy E^k. This leaves us with four variables, two discrete and two continuous. In order to represent G and Σ in the simulation, the continuous variables are sampled along a discrete grid:

$$E^k \rightarrow E_i^k = (i-1)\,\Delta E^k,\; i = 1, N_{E^k} \tag{4.100}$$

$$E \rightarrow E_i = E_0 + (i-1)\,\Delta E,\; i = 1, N_E \tag{4.101}$$

and G and Σ are evaluated on that grid. The start value E_0 of the energy E grid can be chosen so that the grid encompasses the energy range where the Green's functions are significant.

When choosing the grid steps ΔE and ΔE^k of the E^k and E grids, there are two conflicting considerations that require attention: the limited computer memory and the accuracy of the calculations. Many variables like the Green's functions, self-energies and the form factor B are functions of two energy variables, and the arrays representing them will therefore tend to scale inversely proportionally to the square of the step size. To conserve memory, the grid step sizes should be as large as possible. On the other hand, the accuracy and convergence of the algorithm also depend criticially on the step size: the steps should be much smaller than the linewidth of the significant features in the energy spectrum. If this is not the case, the simulation will at best yield questionable results, or may not even converge at all. So we need to find a compromise to accommodate both these issues.

A typical case will serve as an example to provide some useful numbers. Since we are only tracking the Green's functions belonging to one module, the energy range of interest for E encompasses the energies of the functions belonging to that module, i.e. from $|\alpha = 1, E^k = 0\rangle$ to $|\alpha = N, E^k = E^k_{\text{max}}\rangle$. However, it is necessary to set the lower boundary E_0 significantly lower than $E_{|\alpha=1,\mathbf{k}=0\rangle}$ to allow for linewidth broadening of the lowest energy states, various effects of scattering (e.g. LO-phonon scattering) and coherent interactions with lower energy levels in adjacent modules. A similar argument can be applied to the upper boundary of the energy range. In general, an expansion of the energy range by 25 meV on either side is sufficient for our purposes. Assuming $E_N - E_1 \sim 50$ meV, the total energy E range is 200 meV. To provide enough phase space for electrons to scatter and heat up, we usually set E^k_{max}=100 meV. A function with linewidths of about 3 meV requires a grid step of at most 0.3 meV, which translates into almost 700 grid nodes for the E-grid and 350 nodes for the E^k-grid. For a structure with 5 subbands per module, this translates into an array of $5 \times 5 \times 350 \times 700$ or more than 6 million floating point numbers to describe a Green's function. However, the memory requirements are far worse for the form factor B, which also describes the interaction between adjacent modules (i.e. 15 subbands in this example), and would need $15^4 \times 350^2 = 6.2 \times 10^9$ elements. Obviously, this cannot be accommodated and we need to reduce the number of E^k nodes and take advantage of the symmetries in B to further decrease its storage size. The necessary storage space for B can be significantly reduced by switching to a non-uniform E^k grid, which allows for a small ΔE^k for small values of E^k (<10 meV) where B strongly depends on E^k, and larger values of ΔE^k elsewhere. Note that since $E^k \propto k^2$, this choice of grid is more representative of a grid in k-space with a constant step Δk. Finally, many of the interactions represented in B can be

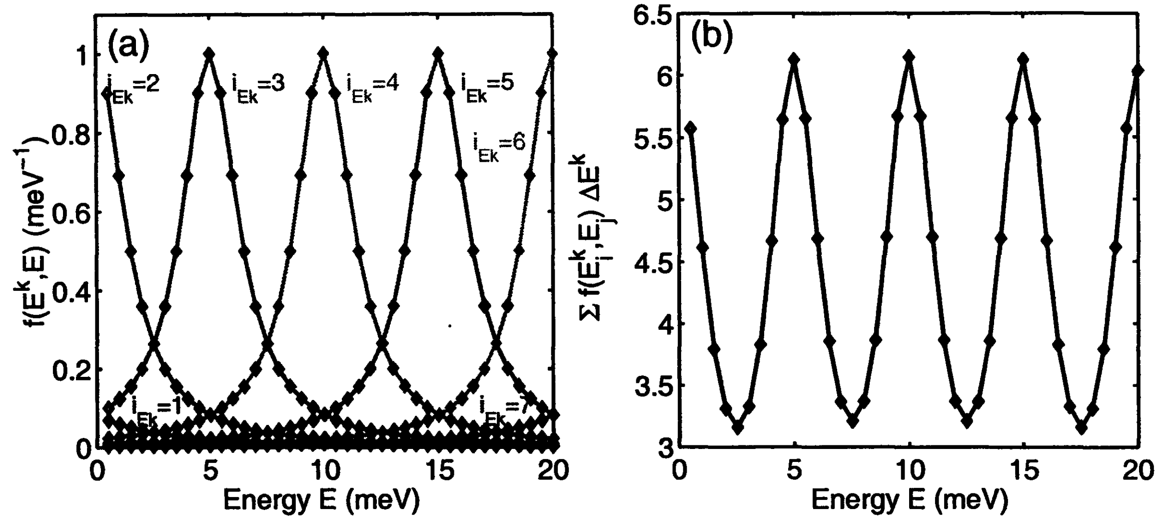

Figure 4-11: *(a) The test function* $f(E^k, E)$*, sampled with* ΔE^k*=5 meV and* ΔE *= 0.5 meV. (b) The integrated function shows peaks and valleys according to the nodes' vicinity to a peak in the sampled function.*

neglected compared to other, competing interactions, which again leads to smaller variables and a large improvement in computation speed.

As mentioned earlier, increasing ΔE^k to be comparable to the linewidth of the Green's functions and self-energies, can lead to errors in the simulation. To illustrate this, consider a function f with a linewidth Γ of 3 meV, and choose $\Delta E = 0.5$ meV $\ll \gamma$ while $\Delta E^k = 5$ mev $\approx \Gamma$, as shown in Figure 4-11(a). Problems arise when we attempt to integrate this function with respect to E^k:

$$\int dE^k f(E^k, E) \to \sum_{E_1^k}^{E_N^k} \Delta E^k f(E_i^k, E), \tag{4.102}$$

which is shown in Figure 4-11(b). Due to undersampling, the numerical integration exhibits peaks and valleys (aliasing), which in this case dip down to half the peak value. To remedy this problem while still maintaining a sparser E^k grid, we can rewrite e.g. Eq. 4.96 in terms of the grid:

$$\int dE^k\, B(E^k, E^{k'})G(E^k, E) = \sum_i \int_{E_i^k}^{E_{i+1}^k} B(E^k, E^{k'})G(E^k, E). \tag{4.103}$$

Each of the terms on the right hand side of Eq. 4.103 relates to one of the terms on the right hand side of Eq. 4.102. Assuming that $B(E^k, E^{k'})$ varies slowly with respect to E^k (which it does), we can take B out of the integration:

$$\sum_i B(E^k, E_i^{k'}) \int_{E_i^{k'}}^{E_{i+1}^{k'}} dE^{k'}\, G(E^{k'}, E) = \sum_i B(E^k, E_i^{k'}) G^{\text{eff}}(E_i^{k'}, E) \Delta E^k \qquad (4.104)$$

where the "effective" Green's function G^{eff} acts as an average over G in the relevant interval, and can easily be evaluated numerically on a denser E^k grid. Using G^{eff} in integrations, it is possible to circumvent the requirement that $\Delta E^k \ll \Gamma$. The minimum step size is now dictated by the changes in B, and $\Delta E^k = 3 - 5$ meV is adequate in most cases.

4.8.5 Main algorithm

Once the scattering matrix elements are set up, we are ready to calculate the Green's functions. Due to the periodic nature of QC structures, the Green's functions G^{mn} describing coherent transport between the subbands of modules m and n are identical to $G^{(m+1)(n+1)}$ of the next modules, save for a shift in energy due to the applied bias voltage:

$$G_{\alpha\beta}^{(m+1)(n+1)}(E^k, E) = G_{\alpha\beta}^{mn}(E^k, E - eFd). \qquad (4.105)$$

If we take into account that G^{ret} is symmetric and $\mathrm{i}G^{<}$ is Hermitian, it is easy to show that we need only keep track of $G^{0(-1)}$ and G^{00} to be able to reconstruct the whole function. We can write for G^{ret}:

$$G^{\text{ret}} = \begin{bmatrix} G^{\text{ret}\,00}(E - E_{\text{bias}}) & [G^{\text{ret}\,0(-1)}(E)]^T & 0 \\ G^{\text{ret}\,0(-1)}(E) & G^{\text{ret}\,00}(E) & [G^{\text{ret}\,0(-1)}(E + E_{\text{bias}})]^T \\ 0 & G^{\text{ret}\,0(-1)}(E + E_{\text{bias}}) & G^{\text{ret}\,00}(E + E_{\text{bias}}) \end{bmatrix}, \qquad (4.106)$$

and for the correlation function:

$$\mathrm{i}G^{<} = \begin{bmatrix} \mathrm{i}G^{<\,00}(E - E_{\text{bias}}) & [\mathrm{i}G^{<\,0(-1)}(E)]^\dagger & 0 \\ \mathrm{i}G^{<\,0(-1)}(E) & \mathrm{i}G^{<\,00}(E) & [\mathrm{i}G^{<\,0(-1)}(E + E_{\text{bias}})]^\dagger \\ 0 & \mathrm{i}G^{<\,0(-1)}(E + E_{\text{bias}}) & \mathrm{i}G^{<\,00}(E + E_{\text{bias}}) \end{bmatrix}. \qquad (4.107)$$

From this perspective, G^{00} describes the intramodule transport, while $G^{0(-1)}$ describes the transport between modules. By using these symmetries, we can significantly speed

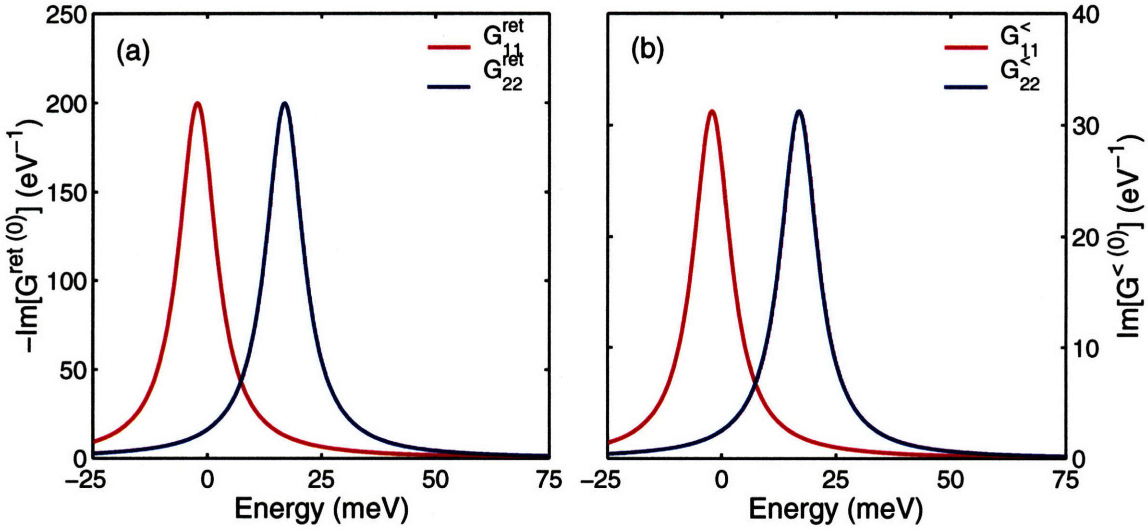

Figure 4-12: *(a) Initial guess for* G^{ret} *(imaginary part) for* E^k*=0 meV for the superlattice. (b) Initial guess for* $G^<$ *(imaginary part) for* E^k*=0 meV.*

up the simulations and reduce the memory requirements to 2/9 or 22%. A similar argument holds true for the self-energies Σ^{ret} and $\Sigma^<$.

The central algorithm to solve the NEGF formalism is based on a very straightforward iteration scheme, as shown in Figure 4-3. As mentioned earlier, all Green's functions and self-energies can be determined from G^{ret} and $G^<$, i.e. if we can find a reasonable initial guess for these functions we can start the iteration. We can use Eq. 4.15 to set up G^{ret}:

$$G^{\mathrm{ret,init}}_{\alpha\alpha}(E^k, E) = \frac{1}{E - E_{\alpha\alpha} - E^k + \mathrm{i}\Gamma}, \tag{4.108}$$

where $E_{\alpha\alpha}$ is the subband energy for level α and Γ is an estimate of the state's lifetime broadening (typically 3-5 meV). Note that the off-diagonal elements of $G^{\mathrm{ret,init}}$ are zero. In keeping with the interpretation of $G^<$ as a generalized density matrix, we can relate $G^<$ to G^{ret} through the occupation probability:

$$G^{<,\mathrm{init}}_{\alpha\alpha}(E^k, E) = -\frac{2\mathrm{i}}{1 + \mathrm{e}^{\frac{E_{\alpha\alpha} + E^k - E_{F,\alpha}}{k_B T}}} G^{\mathrm{ret,init}}_{\alpha\alpha}(E^k, E). \tag{4.109}$$

As can be seen from Figure 4-12, these initial guesses are centered around the semiclassical energy of the state they represent.

With these initial guesses and the previously calculated form factors B, an estimate for the self-energies Σ^{ret} and $\Sigma^{<}$ can be determined, and the Dyson Equation (Eq. 4.15) and Keldysh Equation (Eq. 4.52) in turn provide us with an updated G^{ret} and $G^{<}$. To ensure that the total population density σ remains fixed throughout the simulation, $G^{<}$ is normalized after every iteration:

$$G^{<}_{\alpha\beta}(E^k_i, E_j) \to G^{<}_{\alpha\beta}(E^k_i, E_j) \times \frac{\sigma}{\sigma'}, \tag{4.110}$$

with

$$\sigma' = \frac{m}{\pi\hbar^2} \sum_{\gamma} \int dE^k \int \frac{dE}{2\pi} G^{<}_{\gamma\gamma}(E^k, E). \tag{4.111}$$

To prevent that the large changes in G^{ret} and $G^{<}$ cause instabilities in the algorithm, we can use linear admixing to make the change more gradual. The newly calculated Green's functions are replaced with a weighted sum of the previous and the new estimates. For the Nth iteration:

$$G^N \to cG^N + (1-c)G^{N-1}. \tag{4.112}$$

The parameter c can be adjusted to a higher value as the iteration proceeds to allow for a faster convergence. In our simulations, c is usually chosen in the range $c = 0.5-0.75$. The newly calculated functions G^{ret} and $G^{<}$ can then be used to find a new estimate for the self-energies, and so on. In order to find a self-consistent solution of the transport problem, this process needs to be repeated until the convergence criteria are met, i.e. no value in G^{ret} and $G^{<}$ changes by more than 0.1 % in consecutive iterations:

$$\left| \frac{G^N_{\alpha\beta}(E^k_i, E_j) - G^{N-1}_{\alpha\beta}(E^k_i, E_j)}{G^{N-1}_{\alpha\beta}(E^k_i, E_j)} \right| < 0.001, \quad i = 1 \dots N_{E^k},\ j = 1 \dots N_E. \tag{4.113}$$

Self-consistent solutions found in this manner generally yield a population density σ' that is well within 1 % of σ, which confirms the validity of the solution. Figure 4-13 shows how the Green's functions and the coherent current density converge during the simulation of the superlattice. The scattering current is not calculated until the calculations have converged, and can therefore not be used to monitor the simulation's progress. Note that J_0 remains virtually unchanged after the tenth iteration, while

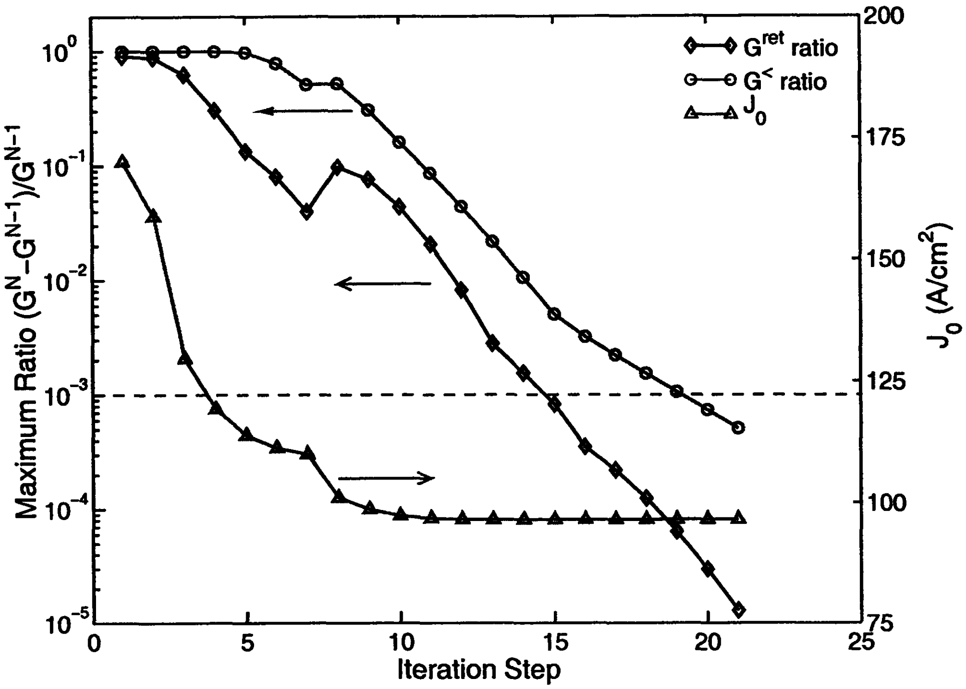

Figure 4-13: *Convergence of* G^{ret} *and* $G^{<}$ *for the simulated superlattice structure. The Green's functions are considered self-consistent when the maximum relative change in any component of* G^{ret} *and* $G^{<}$ *is less than 0.1% (dashed line). Also indicated is the evolution of the coherent current* J_0 *over the course of the simulation.*

the convergence criteria are far from met. This indicates that there is a considerable safety margin and that the obtained current density is a valid simulation result. The implemented convergence criteria are more than adequate to ensure that the simulation results are accurate.

4.8.6 Output

Once the stationary, self-consistent Green's functions and self-energies are found, we can move on to the final stage: the calculation of the current density, electron distribution function and gain. In this section we will examine the calculation results, and discuss the implementation of the gain simulation.

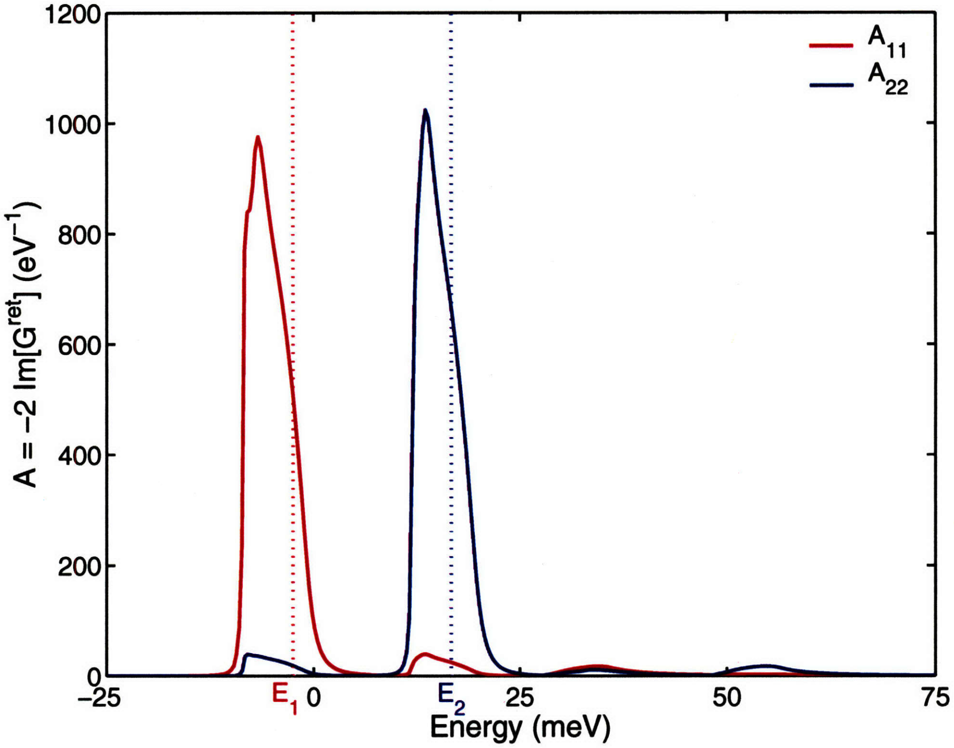

Figure 4-14: *Spectral density functions* $A_{11}(0, E)$ *and* $A_{22}(0, E)$ *as function of energy* E *for the bottom of both subbands* ($E^k = 0$) *of the superlattice.*

Green's functions and self-energies

Although the Green's functions and self-energies do not provide a very obvious and intuitive picture of the transport process they describe, it can nevertheless be very useful to take a closer look at them.

From the imaginary part of G^{ret} we can immediately find the spectral density function (Eq. 4.22), which is shown in Figure 4-14 for the subbands of the simulated superlattice. Note that the peaks of the functions do not match the energy E^ν of the corresponding Wannier function. This mismatch can be attributed to two causes: on one hand, the interaction with the other states, which can cause a positive or negative shift in energy, and on the other hand the real part of G^{ret}, which represents an energy- and momentum dependent energy shift. In general, the visible effect of $\text{Re}(G^{\text{ret}})$ is to cause a fairly uniform shift to lower energies of all Green's functions and self-energies. As the energy differences between different states are almost unaffected, this shift has little physical meaning. The little bumps in each spectral density function are due to coherent interaction, and (as shown later in Fig. 4-16) these disappear when the Hamiltonian is diagonalized.

The density-of-states linewidth is approximately 6 meV for both subbands. It is important to note that the states are only smeared out in energy E space, and not in momentum $\mathbf{k}$ space. We can easily verify this by making a comparison with the semi-classical case. Semi-classically we would expect to see a delta-function at $E = E^{\nu} + E^{k}$, so the total number of states would be unity, as expected:

$$\frac{1}{2\pi}\int dE\,\delta(E - E^{\nu} - E^{k}) = 1. \tag{4.114}$$

In the NEGF formalism the delta function is replaced with the spectral density function:

$$\frac{1}{2\pi}\int dE\,A_{\nu\nu}(E^{k}, E) = \frac{-2}{2\pi}\int dE\,\mathrm{Im}[G^{\mathrm{ret}}_{\nu\nu}(E^{k}, E)] = 1. \tag{4.115}$$

Integrating the functions in Fig. 4-14, we find 0.989 for subband 1 and 0.998 for subband 2, again confirming the convergence and self-consistency of the simulation results.

By examining how G^{ret} changes as function of $\mathbf{k}$ (Figure 4-15), we see that the spectral function grows more symmetric and narrower with increasing momentum. The lopsided appearance of the $E^{k} = 0$ meV curve is due to the proximity of the subband edge, which cuts off the low-energy part of the distribution. The decrease in linewidth can be explained by considering the changes in the (intrasubband) impurity scattering lifetime. Intrasubband elastic scattering at low energies (which roughly corresponds with low E^{k}) is dominated by interactions involving small momentum exchange q (Eq. 4.97), i.e. fast scattering (see Fig. 4-9), while at higher E^{k} the average q, and hence the impurity scattering lifetime, will increase. However, at energies exceeding one LO-phonon energy $\hbar\omega_{\mathrm{LO}}$ above the subband edge, fast intrasubband LO-phonon scattering is allowed and the linewidth increases again. From $E^{k} = 30$ meV to $E^{k} = 40$ meV, the linewidth increases from 2 meV to roughly 4 meV, i.e. an increase of 2 meV, corresponding with a scattering rate of $\tau_{\mathrm{LO}}^{-1} = 2$ meV$/\hbar \approx 3 \times 10^{12}$ s^{-1}.

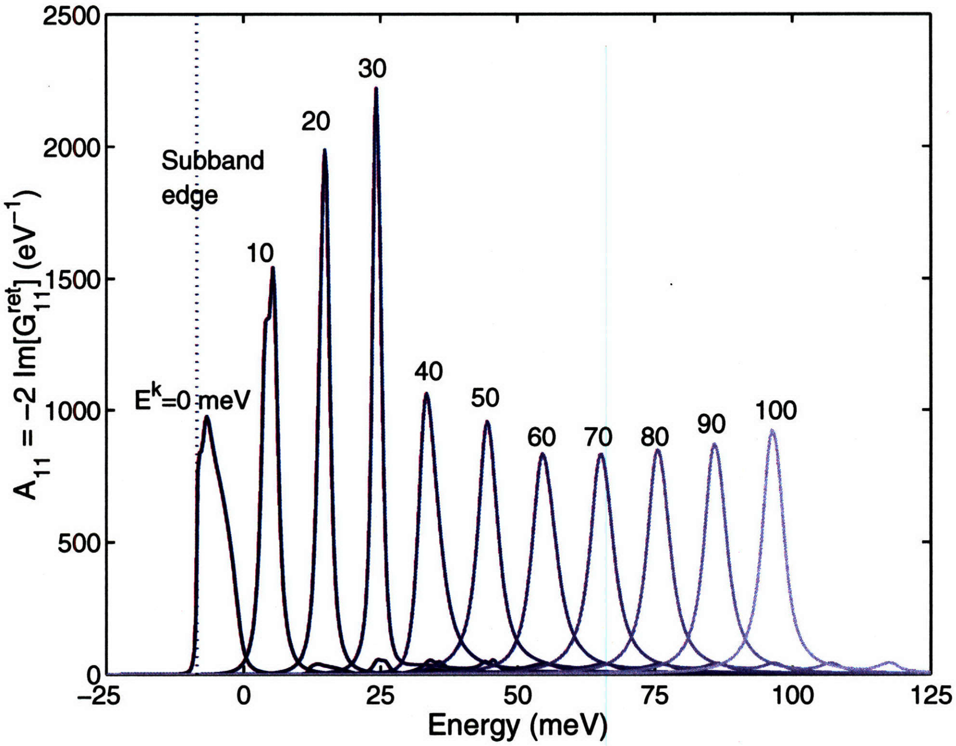

Figure 4-15: *Spectral density function* $A_{11}(E^k, E)$ *as function of energy* E *for different momenta* E^k *in the ground subband of the superlattice. The dotted line indicates the subband edge.*

Figure 4-16(a) shows the calculated $G^<$ for the subbands belonging to the central module and the two adjacent modules. Rather than a spectral density (for G^{ret}) these functions represent the electron distribution function for the states in question.In Figure 4-16(a), the functions for both subbands look very similar in shape (other than a shift in energy) which is not very surprising as they are representing states in anticrossing. The strong interaction between the ground state and the first excited state in the adjacent well can also be seen in the double peak of $G^<_{11}$. The Green's functions belonging to one subband also exhibit a bump around the other subband's energy, which is due to coherent interaction.

Diagonalizing the Hamiltonian and transforming the correlation function according to Eq. 4.83, yields the functions shown in Figure 4-16(b), which now describe the population density in the extended wavefunctions of Fig. 4-5(b). $G^<_{11}$ and $G^<_{22}$ are now very distinct and peak at different energies, in accordance with the relation to

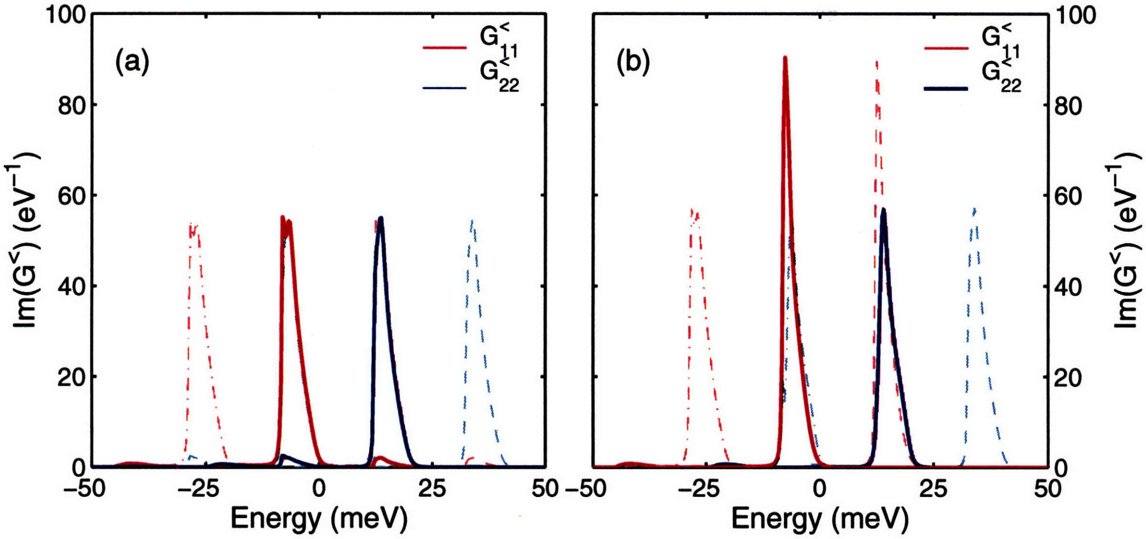

Figure 4-16: *Electron distribution functions* $Im[G^<_{11}(0, E)]$ *and* $Im[G^<_{22}(0, E)]$ *as function of energy* E *for the bottom of both subbands* ($E^k = 0$) *of the superlattice at injection anticrossing. Distributions functions for the adjacent modules are indicated with dashed (previous module - higher energy) and dash-dotted (next module - lower energy) lines. (a) Wannier basis. (b) Transformed basis (extended wavefunctions).*

their anticrossing wavefunctions. Also note that the "coherence bump" is no longer present in the transformed $G^<$.

By integrating $G^<$ over E we can find the density matrix $\rho(E^k)$ (Eq. 4.27) and the semi-classical electron distribution function $f(E^k)$, which yield the subband populations. Table 4.2 shows that the ground state is slightly more populated than the excited level, which indicates that transport is affected by dephasing. In the diagonalized basis, the population difference does not disappear as it would for extended wavefunctions in the MCDM simulations. The density matrix formalism retains no information about the energy spectrum, and when one state anticrosses with another, their populations are redistributed regardless of their energy spectra. In the absence of dephasing, anticrossing two states with the same energy will result in their populations being distributed evenly across the interacting states, whatever their energies are. Because the DM model does not keep track of energy, the coherent interaction it describes only enforces conservation of momentum. In the NEGF formalism however, the states' coherent interaction conserves both momentum ($\mathbf{k}$) and energy (E).

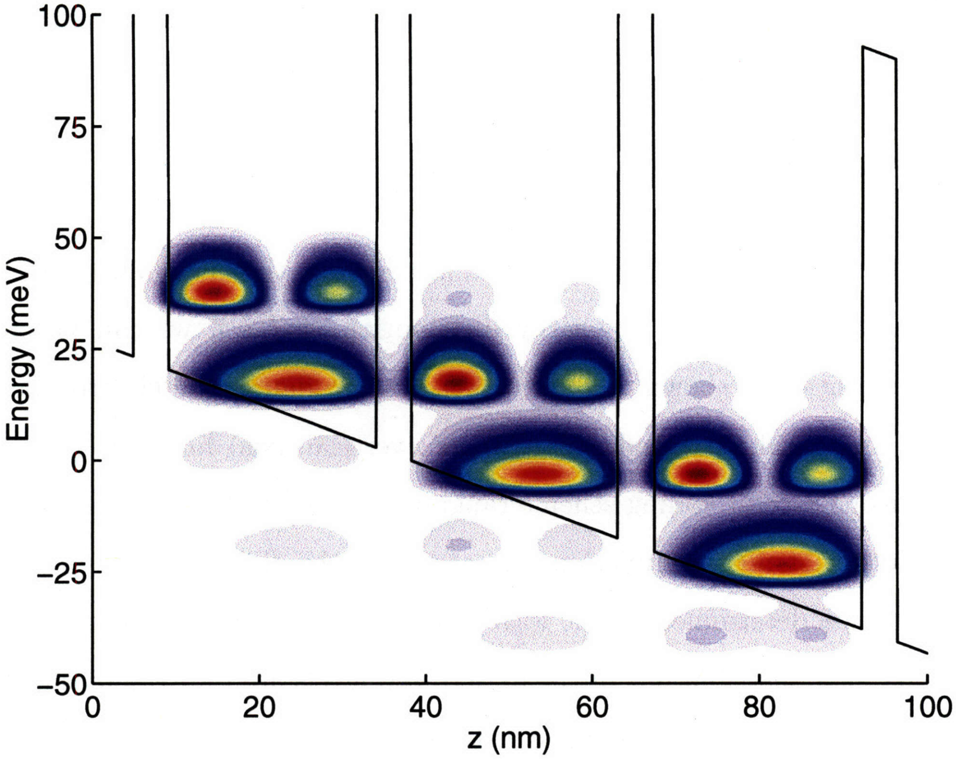

Figure 4-17: *Electron distribution in the superlattice at anticrossing, assuming diagonalized wavefunctions. Indicated are the electron density in real space (horizontal axis) and energy space (vertical axis) for both subbands, using a color coding scheme ranging from red (high density) to blue (low density).*

The anticrossing of two (unperturbed) states with the same semi-classical energy and thermalized populations will give rise to anticrossed states where the lower-energy state has a higher population density than the excited state. The density of states of the ground state is higher at low energies, and since there are more electrons available at those energies (due to the original thermalized distributions) this state will have a higher population density.

Subband (Wannier)	Pop. dens. ($\times 10^9$ cm^{-2})	Subband (diag.)	Pop. dens. ($\times 10^9$ cm^{-2})
1	5.4	1	5.6
2	4.6	2	4.4

Table 4.2: *Subband populations for the superlattice at injection anticrossing, using a Wannier basis and the basis wavefunctions of the diagonalized Hamiltonian.*

Since the relation between $\mathbf{k}$ and E is no longer unique, it is instructive to redefine the electron distribution function as a function of E instead of E^k:

$$f(E) = -\mathrm{i}\sum_{\mathbf{k}} \frac{1}{2\pi} G^{<}(E^k, E). \tag{4.116}$$

Combining the spatial electron distribution, provided by the diagonalized basis wavefunctions, and the transformed energy distribution of $f(E)$, we can visualize the population densities as shown in Figure 4-17. In the figure, the electron density in real space (horizontal axis) and energy space (vertical axis) for both subbands are illustrated, using a color coding scheme ranging from red (high density) to blue (low density). Note that no wavefunctions are shown in this figure, although the probability distribution determined by the wavefunctions is clearly reflected in the electron distributions. The distributions of the anticrossed levels are indistinguishable on the figure and can be thought of as a single distribution spread over two subbands. While most electrons are concentrated near the bottom of the subbands, some electrons are also visible at roughly $\hbar\omega_{\mathrm{LO}}$ below the main distributions. It is not immediately clear whether this reflects a real physical phenomenon or is merely an artifact of the simulation.

To get an idea about scattering rates, we can take a closer look at the self-energies Σ^{ret} and $\Sigma^{<}$, which are shown in Fig. 4-18. The imaginary part of the retarded self-energy is directly related to the raw scattering rate Γ, and allows us to estimate the level's lifetime. The two major contributions to scattering, elastic scattering in the form of electron-impurity scattering and inelastic scattering in the shape of LO-phonon scattering, are both easy to recognize when examining Fig. 4-18(a). We will focus on $\Sigma^{\mathrm{ret}}_{11}$, describing the scattering in the ground state n=1. The peak around $E = E_1$ clearly describes an elastic scattering process, i.e. intrasubband impurity scattering and interface roughness scattering. Note that the maximum does not exactly coincide with E_1, similar to G^{ret}. Intersubband elastic scattering causes another very small feature around $E = E_2$. The small bump around E=-42 meV is due to resonant LO-phonon absorption, which is of little significance due to the low lattice temperature ($T_{\mathrm{latt}} = 25K$). The LO-phonon emission peaks above E=30 meV are

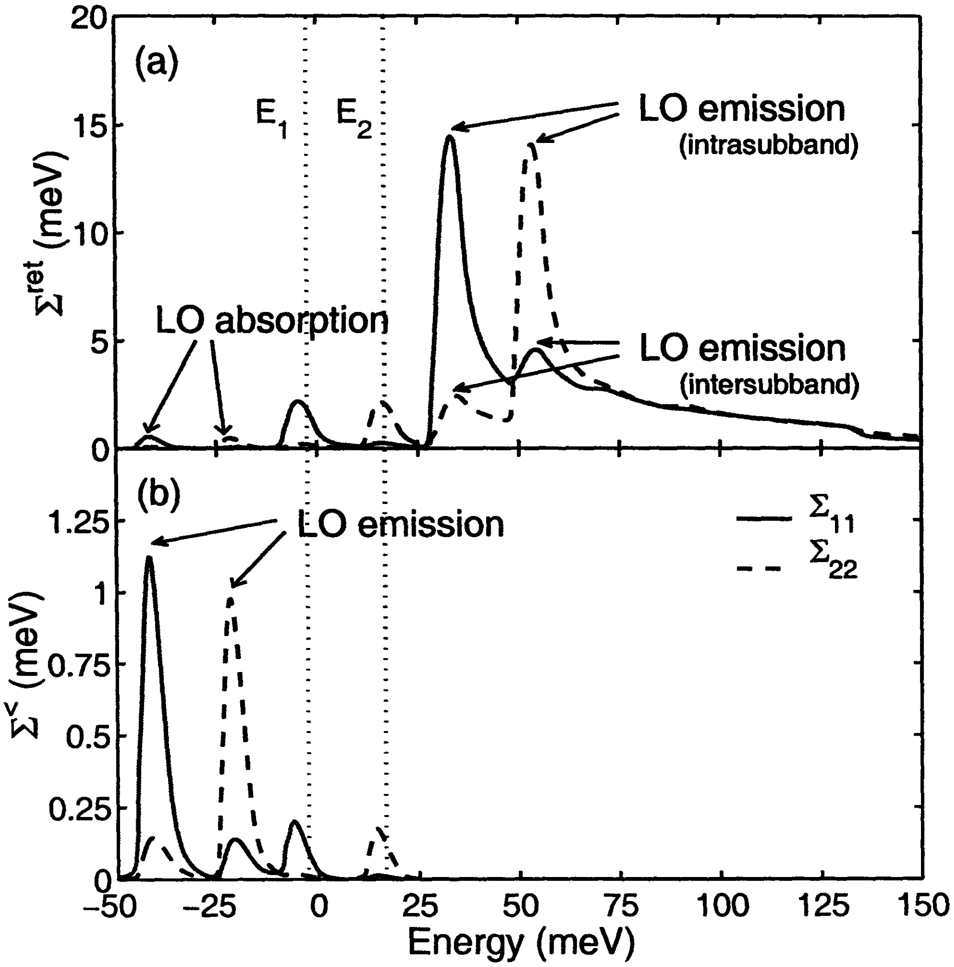

Figure 4-18: *(a) Imaginary part of* Σ^{ret} *for both subbands of the superlattice structure at* $E^k = 0$ *meV. Also indicated are the semi-classical energies* E_1 *and* E_2. *(b) Imaginary part of* $\Sigma^{<}$ *for both subbands of the superlattice structure at* $E^k = 0$ *meV.*

much more prominent. Due to the small overlap between the ground (Wannier) states in adjacent wells, LO-phonon emission from n=1 is only possible for an intrasubband transition ($E > E_1 + \hbar\omega_{\text{LO}}$) or an intersubband transition to n=2 ($E > E_2 + \hbar\omega_{\text{LO}}$). As can be seen in Fig. 4-18(a), the scattering rates due to these transitions decrease quickly as E increases and we get farther away from resonance. We can find the scattering lifetime for the intrasubband LO-phonon peak around E=30 meV (Γ=30 meV) from $\tau_{\text{LO}} = \pi\hbar/\Gamma$=0.28 ps, i.e. the resonant LO-phonon scattering lifetime. Note that this peak describes the transition of an electron with E^k=0 meV (hence $\mathbf{k} \approx 0$) and energy $E \approx 30$ meV to other states near the subband edge ($\mathbf{k}' \approx 0$)

which have a substantial density of states near E = 30 meV - $\hbar\omega_{\rm LO}$ = -6 meV (as shown in Fig. 4-14). In other words, due to the extended energy spectrum, the NEGF formalism allows for (resonant) intrasubband LO-phonon scattering (with very small q) between states with similar momentum $\mathbf{k}$. However, this does not imply that the actual electron flow due to this process will be significant, to get a high electron flux there needs to be a substantial population density (found from $G^<$ in figure 4-16) at that energy (and momentum) as well as a high intrinsic scattering rate. This description is in sharp contrast to the semi-classical picture, where intrasubband LO-phonon scattering implies a much larger $q > \sqrt{2m\omega_{\rm LO}/\hbar} \approx 2.5 \times 10^8$ m^{-1}, and therefore a much smaller scattering rate.

As explained in section 4.4, the lesser self-energy $\Sigma^<$ measures the rate at which electrons scatter into a state. Analogous to $\Sigma^{\rm ret}$, we can distinguish the contributions from elastic scattering (around the semi-classical energies) and intra- and intersubband LO-phonon emission, at $E = E_1 - \hbar\omega_{\rm LO}$ for $\Sigma^<_{11}$. Note that the LO-phonon emission peaks (electrons scattering into the state due to LO-phonon emission out of another state) occur at energies lower than the semi-classical energy, in contrast to $\Sigma^{\rm ret}$. This can also be seen immediately in the expression for $\Sigma^{\rm ret,LO}$ and $\Sigma^{<,\rm LO}$ (Eqs. 4.39 and 4.40), where the terms for LO-phonon emission are proportional to $G^{\rm ret}(E-\hbar\omega_{\rm LO})$ and $G^<(E+\hbar\omega_{\rm LO})$, respectively. Resonant LO-phonon emission (with small q) from states around the subband edge ($\mathbf{k} \approx 0$, with population densities centered around $E = E_1$) is responsible for the LO-phonon emission peak, and causes an in-flux of electrons around $E = E_1 - \hbar\omega_{\rm LO}$.

Gain Simulation

The calculation of the (small-signal) gain is one of most important aspects, if not the most important aspect of a (QCL) device simulation. The optical characteristics of the investigated device, its emission energy, linewidth and peak gain will to a large extent determine its success or failure as a laser. The Monte Carlo simulations, both semi-classical and density matrix, contain no direct information about the linewidth (other than inferred from scattering times), and assume a Lorentzian

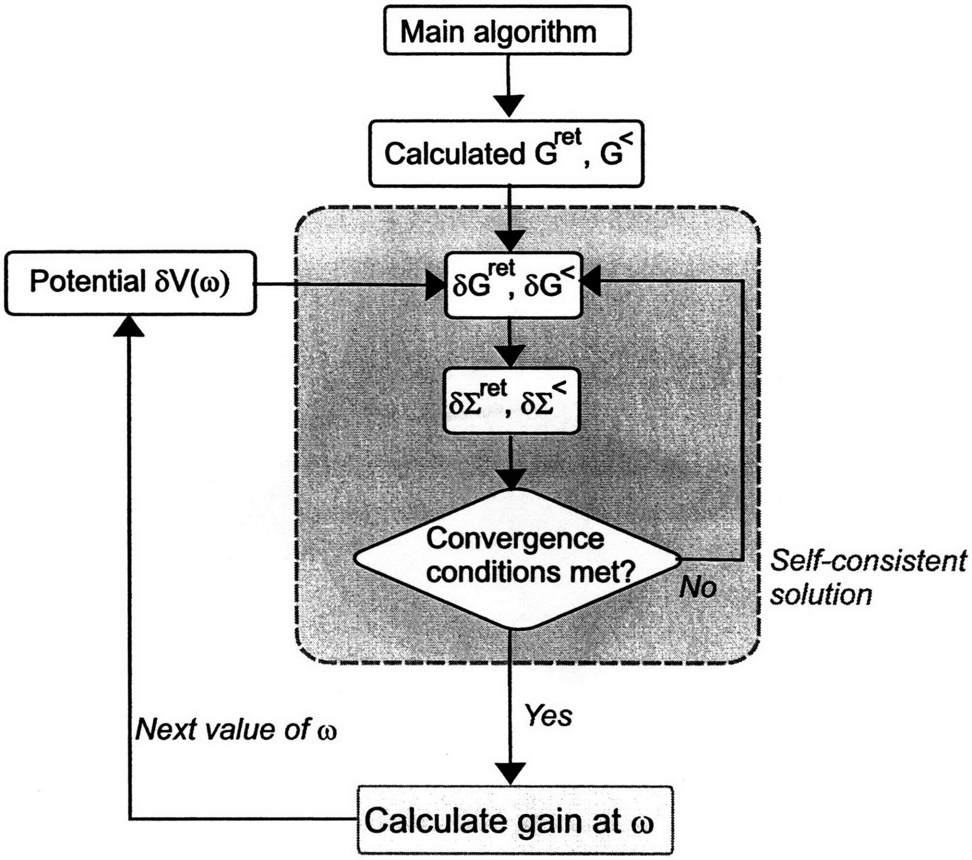

Figure 4-19: *Flow chart for the calculation of the gain spectra starting from the stationary Green's functions.*

lineshape centered around the targeted frequency to obtain a peak gain value. The correct values for the spontaneous emission linewidth, the emission frequency, the oscillator strength and even the population inversion are debatable in many cases, and lead to much uncertainty in the interpretation of the simulation results. The exact wavefunctions and associated populations, and hence the oscillator strength and population inversion, involved in the optical transition are approximated differently in every simulation, and even within a single simulation different oscillator strengths and populations may be used in the gain calculations depending on how strong the role of resonant tunneling is thought to be in the injection of electrons into the upper radiative state. For exaple, in the semiclassical simulation of FL175C, we may model the gain as primarily due to the 5-4 transition, or include the broadened 1'-4 transition and determine the maximum gain due to contributions from both broadened

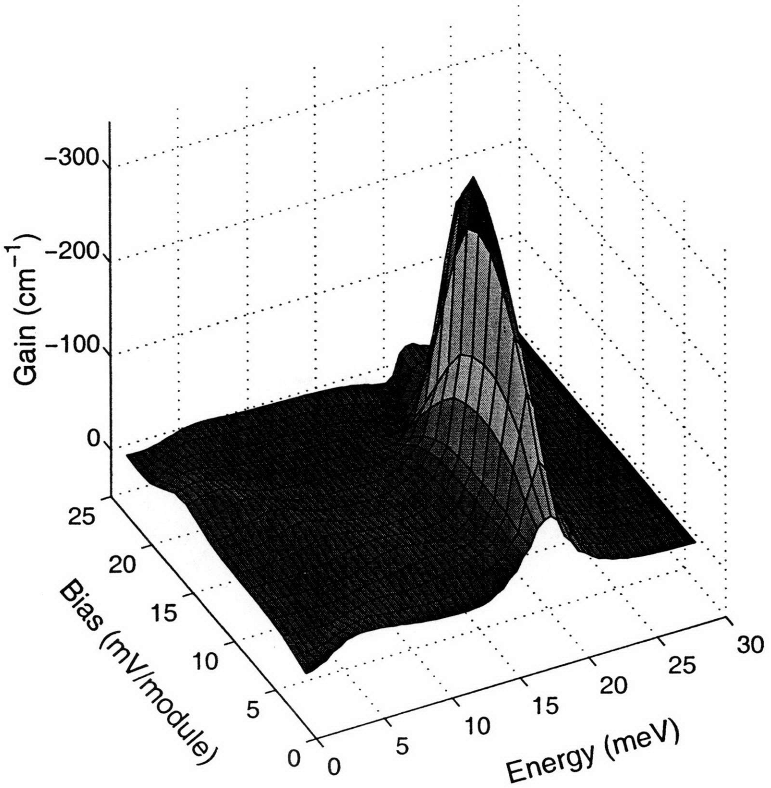

Figure 4-20: *Gain versus emission energy and bias voltage for the superlattice. The gain axis is reversed for clarity.*

transitions.

On the other hand, the implicit description of the energy spectra in the NEGF formalism allows for a very inclusive calculation of the gain spectrum, yielding information about the peak gain, frequency, linewidth and lineshape. This added information removes the need for the assumptions made for the MC gain calculations, providing a more robust model for the gain.

As explained in section 4.6, the optical gain is essentially calculated as a small-signal perturbation to the large-signal Green's functions we obtained from the main simulation. To resolve the interdependence of the small-signal changes in the Green's functions and self-energies, another iterative scheme is required, similar to the one used for the large-signal functions. However, there is one significant difference: the time-dependence of the perturbing optical potential $\delta V(\omega)$ lifts the symmetry of G^{ret} and Σ^{ret}, and the Hermitian conjugation symmetry of $\mathrm{i}G^{<}$ and $\mathrm{i}\Sigma^{<}$. For a superlattice or QC structure, the Green's functions belonging to the states in one module are still identical to the Green's functions in another well, except for a translation in energy. We can write for G^{ret}:

$$\delta G^{\mathrm{ret}} = \begin{bmatrix} \delta G^{\mathrm{ret}\,00}(E - E_{\mathrm{bias}}) & \delta G^{\mathrm{ret}\,01}(E - E_{\mathrm{bias}}) & 0 \\ \delta G^{\mathrm{ret}\,0(-1)}(E) & \delta G^{\mathrm{ret}\,00}(E) & \delta G^{\mathrm{ret}\,01}(E) \\ 0 & \delta G^{\mathrm{ret}\,0(-1)}(E + E_{\mathrm{bias}}) & \delta G^{\mathrm{ret}\,00}(E + E_{\mathrm{bias}}) \end{bmatrix}, \quad (4.117)$$

and for the correlation function:

$$\mathrm{i}\delta G^{<} = \begin{bmatrix} \mathrm{i}\delta G^{<\,00}(E - E_{\mathrm{bias}}) & \mathrm{i}\delta G^{<\,01}(E - E_{\mathrm{bias}}) & 0 \\ \mathrm{i}\delta G^{<\,0(-1)}(E) & \mathrm{i}\delta G^{<\,00}(E) & \mathrm{i}\delta G^{<\,01}(E) \\ 0 & \mathrm{i}\delta G^{<\,0(-1)}(E + E_{\mathrm{bias}}) & \mathrm{i}\delta G^{<\,00}(E + E_{\mathrm{bias}}) \end{bmatrix}. \quad (4.118)$$

Compared to the expressions in Eqs. 4.106 and 4.107, now we also need to keep track of the interactions with the "next" module δG^{01} besides δG^{00} and $\delta G^{0(-1)}$ to get a full picture of all interactions.

Since the small-signal functions are generally a function of the excitation frequency ω, the iterative process has to be repeated for all desired ω in the target frequency range. The calculation scheme is shown in Fig. 4.8.6. Similar to the main algorithm, new functions δG^{eff} are defined to avoid aliasing effects due to a coarse **k** grid (see Eq. 4.104.

Gain spectra for several different biases are shown in Figures 4-20 and 4-21, for emission energies between 1 and 30 meV. Also shown are the extended wavefunctions corresponding to each bias, found by transforming the original Wannier functions (Eq. 4.83). For low biases, up to 16 meV/module, the gain spectrum exhibits a clear negative peak around 19 meV, which corresponds with the intersubband energy separation between n=1 and n=2. For these biases, most electrons are stuck in the ground state.

Closer to the injection anticrossing, the ground state n=1 and the excited state in the next well n=2" start interacting, dropping n=1 to a lower energy while pushing the energy of n=2" (and equivalently n=2) up, which results in a higher E_{21}. This blue shift in the loss can be seen in the figure for biases higher than 16 meV/module.

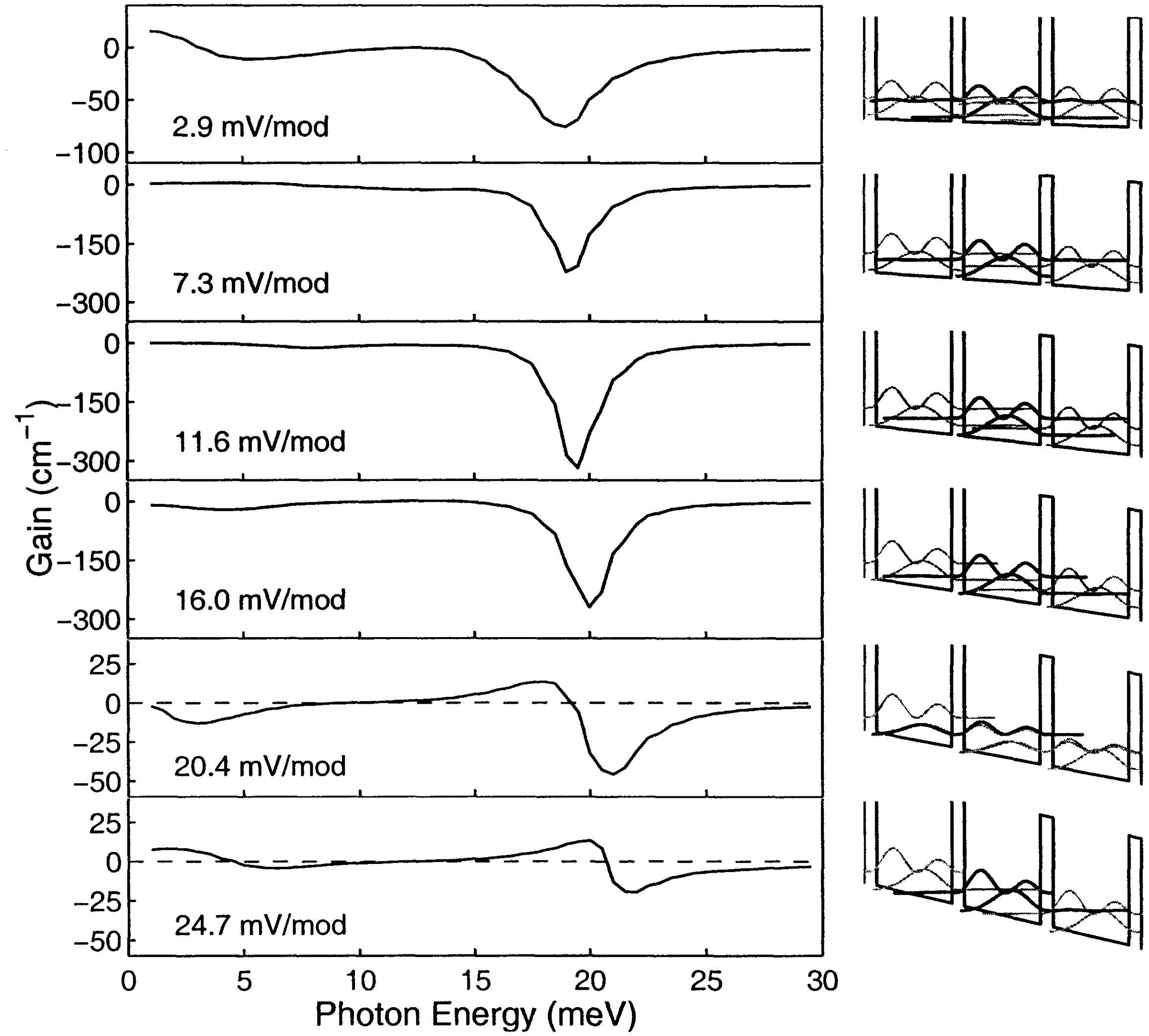

Figure 4-21: *Gain spectra for different applied electric fields for the superlattice structure. Extended wavefunctions are shown on the right-hand side.*

More electrons are leaking into the excited state, making the population inversion Δn_{21} less negative and decreasing the absolute value of the peak loss. At the injection anticrossing (20.4 mV/module), we can see a peak loss of 50 cm^{-1} at $\hbar\omega$=21 meV, and a small gain of approximately 10 cm^{-1} at $\hbar\omega$=18 meV. This can be understood when we consider that the loss peak is due to the optical interaction between the lowest-energy state of the ground doublet and the highest-energy state of the excited doublet (wavefunctions indicated in red), with a corresponding negative population inversion. However, the population inversion between the two other states in both doublets is of equal magnitude, but positive, and leads to a positive gain value at a lower energy. If we consider the doublet as a single, highly interacting "state", this picture is very similar to Bloch gain, which also shows a double loss-gain peak. However, due to dephasing and wavefunction localization the population density localized at the bottom of the well is slightly higher than the population concentrated at higher energy, which accounts for the fact that the peak loss is higher than the peak gain.

Chapter 5

Simulation Results

To compare the different approaches to the modeling of electron transport in QCLs, the density matrix model and a semi-classical Monte Carlo simulation as well as a Green's function simulation with and without momentum-dependent scattering (referred to as "full" and "simple" NEGF) were used to calculate current densities, populations and electron temperatures for several different QC designs. We will focus on three representative examples, for which experimental results were published, and use the semi-classical MC simulation to investigate the importance of e-imp scattering in QCLs. The Monte Carlo simulations include semi-classical electron-phonon (acoustic and LO), electron-impurity (e-imp) and electron-electron (e-e) scattering, and a nonequilibrium, multisubband screening model for e-imp and e-e interactions. The NEGF simulations include electron-LO-phonon, electron-impurity, electron-interface-roughness and electron-electron scattering (only for the simulation with momentum dependent scattering) and the same screening model as the MC simulations. Interface roughness scattering with roughness Δ=2.825 Å(one monolayer) and correlation length Λ=5 nm was included in the NEGF calculations, but its influence on the simulation results was negligible. The only phenomenological parameter is the pure dephasing time T_2, used in the DM-MC calculations. In all simulations, except for Fig. 5-12, the lattice temperature T_{latt} was assumed to be 25 K.

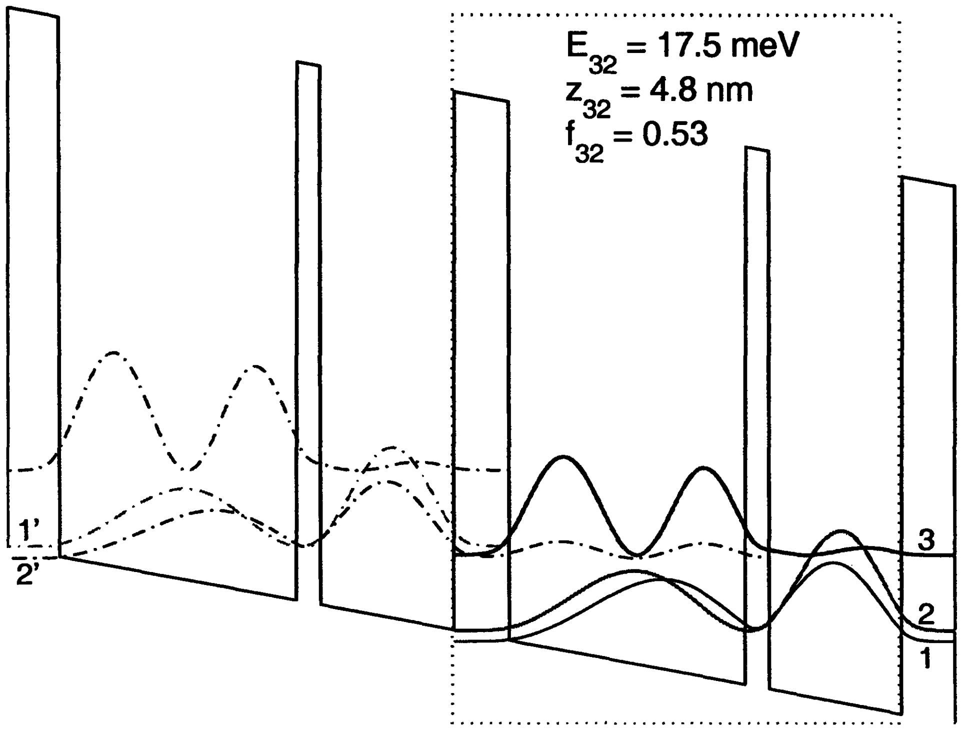

Figure 5-1: *Band structure of T65. The device consists of GaAs/*$\mathrm{Al_{0.15}Ga_{0.85}As}$ *layers with thicknesses (nm)* **5.5**/*23.4*/**2.4**/*13.2 (barriers in boldface, wells in plain text) and a doping concentration* n=*1.4*$\times 10^{16}$ *cm*$^{-3}$ *in the 13.2-nm wide well, resulting in a sheet density of 1.85*$\times 10^{10}$ *cm*$^{-2}$ *per module.*

5.1 T65

The first investigated device is a simple double-quantum-well structure [56] (T65), whose conduction band diagram and wave functions are reproduced here in Fig. 5-1. The active region consists of a 23.4-nm wide well where the intrawell radiative transition takes place primarily between levels n=3 and n=2. A narrower 13.2 nm collector well and its associated subband n=1 are used to collect electrons and inject them into the next module. The narrower well is doped to provide an electron density of 1.85 x 10^{10} cm^{-2} per module.

In experiments, the current characteristics of this device were nearly independent of temperature for a lattice temperature between 5 and 77 K, while the electroluminescence (proportional to τ_3) dropped by a factor of 1.5. No gain or superlinear power-current ($L-I$) relations were ever observed. This is a strong indication that transport through the thick injector barrier is limited by incoherent tunneling between the injector state $n=1'$ and the upper radiative level $n=3$ ($\Delta_{1'3} \approx 0.8$ meV). Furthermore, magnetotunneling spectroscopy revealed clear evidence of an anticrossing gap of $\Delta_{21} \approx 2$ meV between levels 1 and 2, validating our hybrid model of using a semi-classical model for intramodule transport.

The DM simulations included a phenomenological dephasing time of 0.5 ps, which is consistent with the measured spontaneous emission linewidth Δf of 2 meV (0.5 THz, and $\Delta f \approx h/\pi T_2$).

Figure 5-2 shows the calculated and measured current density for a lattice temperature of T_{latt}=25 K. The DM calculations show a peak current density of 116 A/cm^{-2} that corresponds well with the experiments (123 A/cm^{-2}), while the semi-classical model overestimated the peak current density (186 A/cm^{-2}). Both Green's function simulations produced an almost identical peak current density of 130 A/cm^{-2}.

In all simulations there is a bump in the $I-V$ characteristic around 20 mV/module, close to the anticrossing of $n=2'$ and $n=3$. This feature is most evident for the simple NEGF model, and it predicts rise to a small region of negative differential resistance (NDR) between 17 and 21 mV/module. However, there is no evidence of this transition in the experimental results. Overall, the experimental current density seems to be reproduced most faithfully both qualitatively and quantitatively by the DM simulation. The sharp features in both the current density and electron temperature (Fig. 5-3), which are clearly visible in the semi-classical and simple NEGF

Table 5.1: *Calculated subband energy, temperature, and population density of T65 at injection anticrossing (T_{latt}=25 K) for the semi-classical, density matrix and full NEGF simulations.*

n	E (meV)	T_{el} (K)			Pop. (10^{10} cm^{-2})		
		semi-cl.	DM	Full NEGF	semi-cl.	DM	Full NEGF
1	0	50	86	82	0.53	0.91	0.69
2	3.9	85	95	82	0.37	0.58	0.68
3	25.0	64	62	63	0.95	0.42	0.46

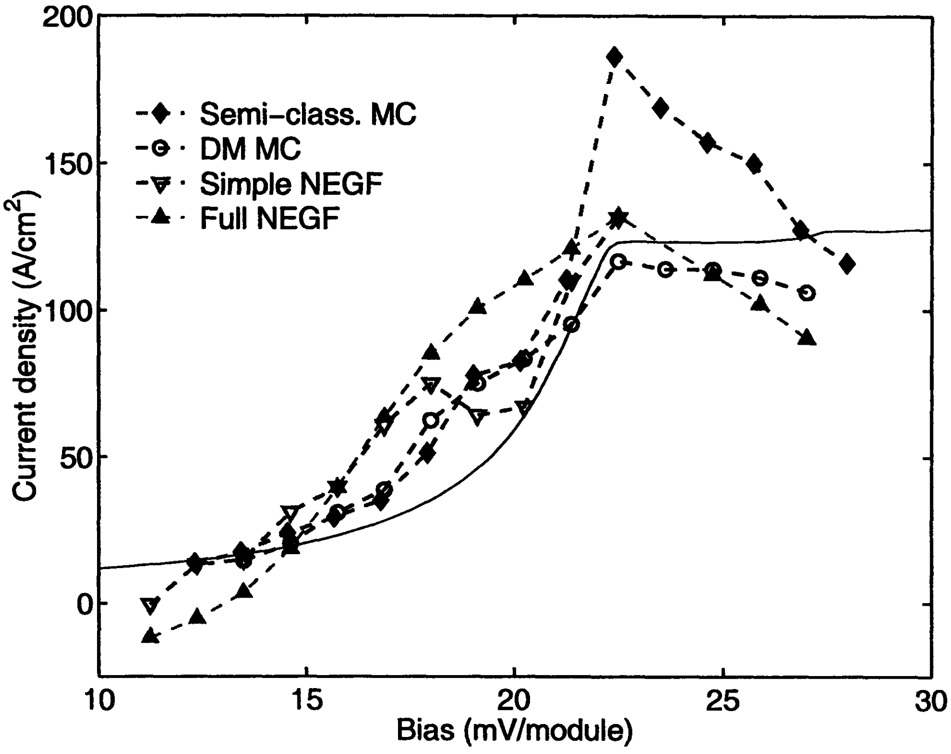

Figure 5-2: *Current density for T65 for a range of biases for the MC and NEGF simulations. The measured current density (solid line) was adjusted to take into account a parasitic series resistance of 1.5 Ω.*

calculations, are smoothed out in the DM and full NEGF results.

Figure 5-4 shows a graphical representation of the spatial and energetic distribution of the electrons in the structure at injection anticrossing, obtained from the full NEGF results. There is no clear distinction between the populations of the different subbands that make up the triplet. The highest electron density is concentrated in the narrow well, suggesting that the injection barrier inhibits transport and causes a pile-up of electrons in the injector state. Fig. 5-5 illustrates the various results for the electron distribution functions obtained from the different simulations at injection anticrossing. The distributions below E^k=12 meV can be described as thermal for all models except for the simple NEGF, due to its absence of e-e scattering. The distribution function for n=3 exhibits a downward kink around E^k=15 meV for all simulations, which corresponds to the onset of (thermally activated) LO-phonon scattering from n=3 to the lower energy subbands. Note that the absence of e-e scattering

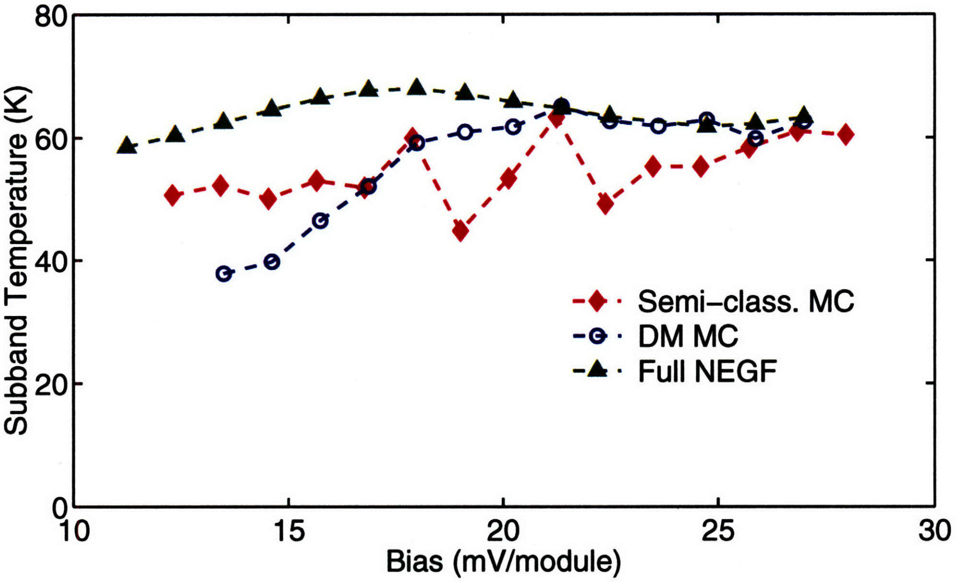

Figure 5-3: *Electron temperature for n=3 for the MC and full NEGF simulation (lattice temperature* T_{latt}*=25 K). The simple NEGF simulation is not represented due to its non-thermal distributions.*

in the simple NEGF model is also reflected in the small increase of the distributions for n=2,3 around E^k=20 meV (due to the aforementioned e-LO scattering from n=3) while this feature is smoothed out in the full NEGF model. A second kink can be seen in all distribution functions at $E^k = \hbar\omega_{\mathrm{LO}}$ due to intrasubband LO-phonon scattering. The presence of these kinks in all simulations suggests that intrasubband e-e scattering is not fast enough to efficiently redistribute electrons within n=3 as the "hot" electrons with energies exceeding $E^k \approx$15 meV are scattered with a lifetime $\tau_{\mathrm{LO}} \approx$0.3 ps. While the low-energy part of the distribution can be described as thermal, the high-energy tail is decidedly non-thermal. This means that simple, back-of-the-envelope calculations to determine the e-LO scattering rate based on a thermal distribution may be substantially overestimating its importance.

Most important is the absence of population inversion (Fig. 5-6), and hence gain, in the DM calculation, which is in agreement with the experimental results. A more ambiguous picture is presented by the full NEGF simulations (Figure 5-7 for the design bias), which shows an absorption peak of 26 cm^{-1} around 21 meV and a small gain peak of 9 cm^{-1} at 18 meV. Below 10 meV the "intraminiband" absorption

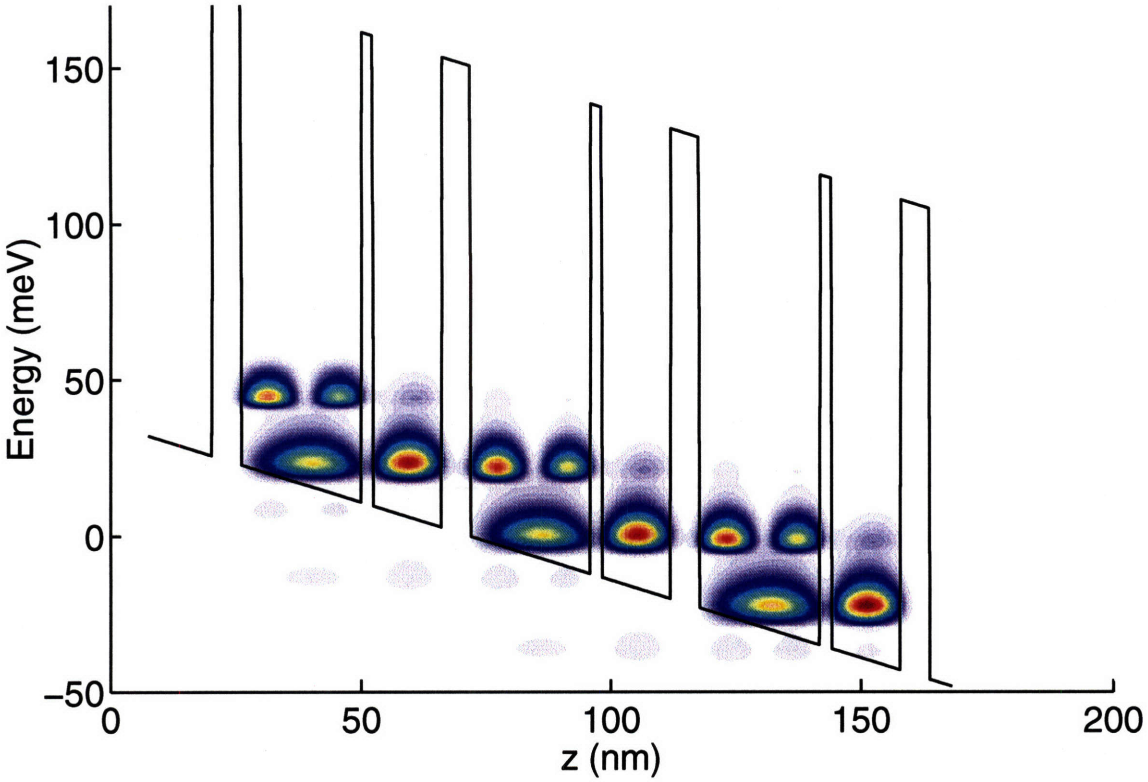

Figure 5-4: *Electron distribution in T65 at injection anticrossing, assuming diagonalized wavefunctions. Indicated are the electron density in real space (horizontal axis) and energy space (vertical axis) for both subbands, using a color coding scheme ranging from red (high density) to blue (low density).*

between the subbands of the triplet n=2,1,3" starts to dominate. The combination of the gain peak at lower energies and absorption peak at higher energies is similar to the gain described in the discussion of the two-level superlattice, with the distinction that the earlier doublet is replaced with a triplet. Both peaks have a linewidth of about 3 meV, which is slightly higher than the observed Δf=2 meV. On the other hand, the semi-classical results indicate a maximum population inversion of $n_3 - n_2 \approx 6 \times 10^9$ cm^{-2} (corresponding to a $\Delta n_{3\mathrm{D}} \approx 1.4 \times 10^{15}$ cm^{-3}), and a predicted gain of 180 cm^{-1} for a spontaneous emission linewidth Δf=2 meV. Such a high level of gain would have made lasing quite easy to achieve. However, no lasing was observed from T65 and similar structures, even embedded in metal-metal waveguides. [38] The DM and NEGF analyses indicate that it is likely that electrons are "stuck" behind

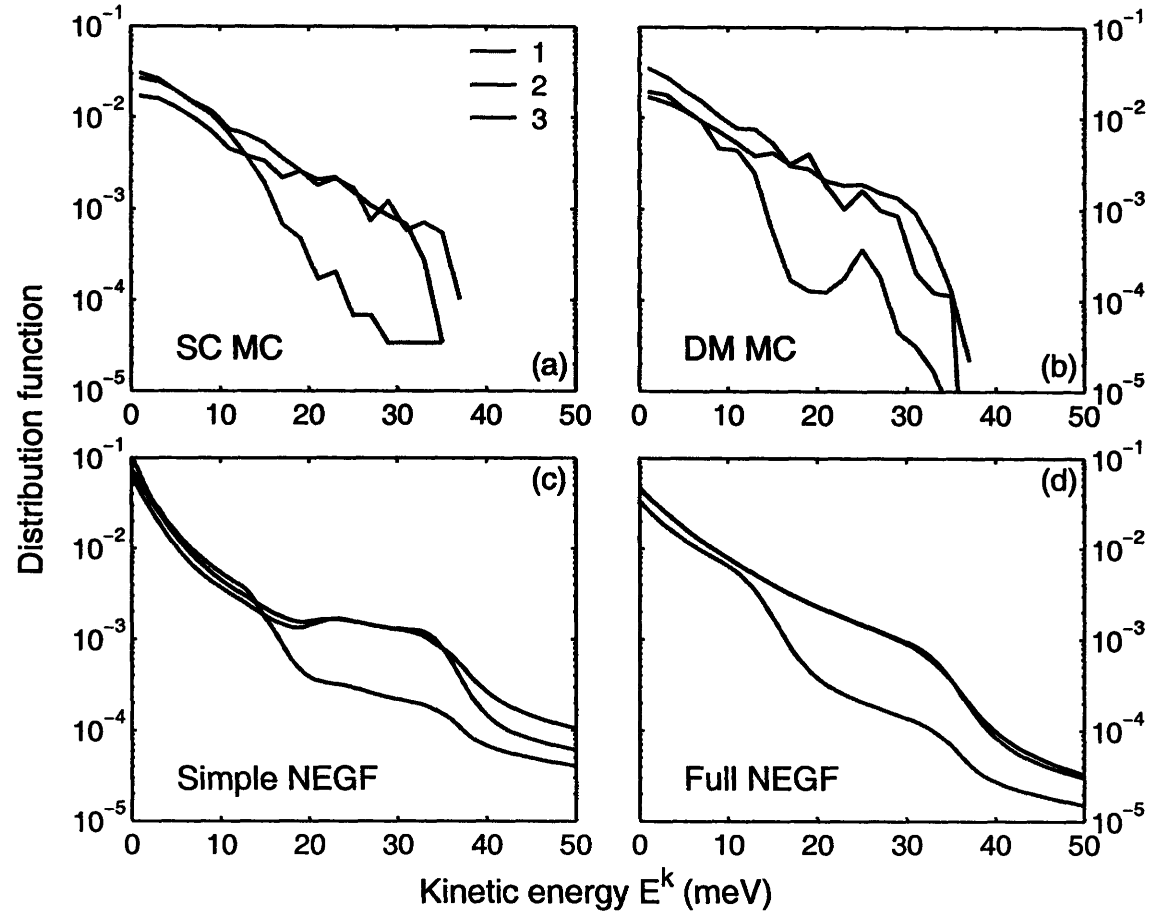

Figure 5-5: *Electron distribution functions of the three lowest-energy subbands in T65 at injection anticrossing. The NEGF distribution functions are for the Wannier states and are not diagonalized.*

the injection barriers, residing mostly in levels 1′ and 2′ instead of equally populating 1′ and 3 as the semi-classical model predicted.

5.2 FL177C-M5

The second QC device is a 3.2 THz QCL, [57] labeled FL177C-M5, which operated up to ~ 130 K in pulsed mode, and whose band diagram and wavefunctions are reproduced in Fig. 5-8. Electrons injected into the upper radiative state n=5 make a radiative transition down to n=4, which is in anticrossing with the excited state n=3 in the wide well. The energy separation between n=3 and the injector/collector

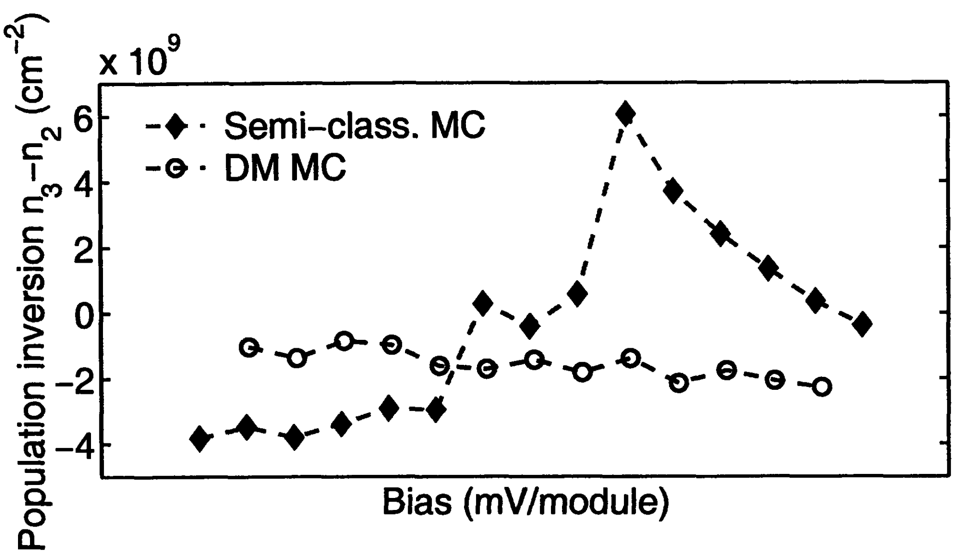

Figure 5-6: *Population inversion $n_3 - n_2$ for the MC simulations for a range of biases.*

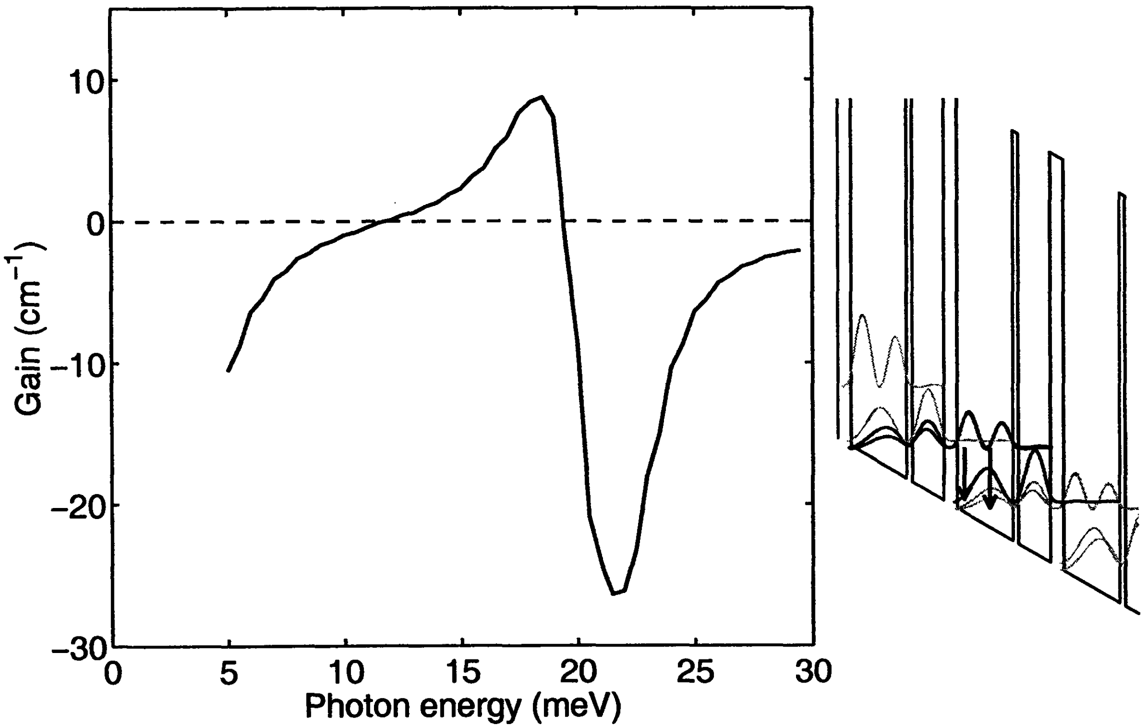

Figure 5-7: *Calculated gain spectrum for T65 at injection anticrossing calculated with the full NEGF simulation. Also indicated are the diagonalized, extended wavefunctions at this bias and the transitions corresponding to the peaks in the gain spectrum.*

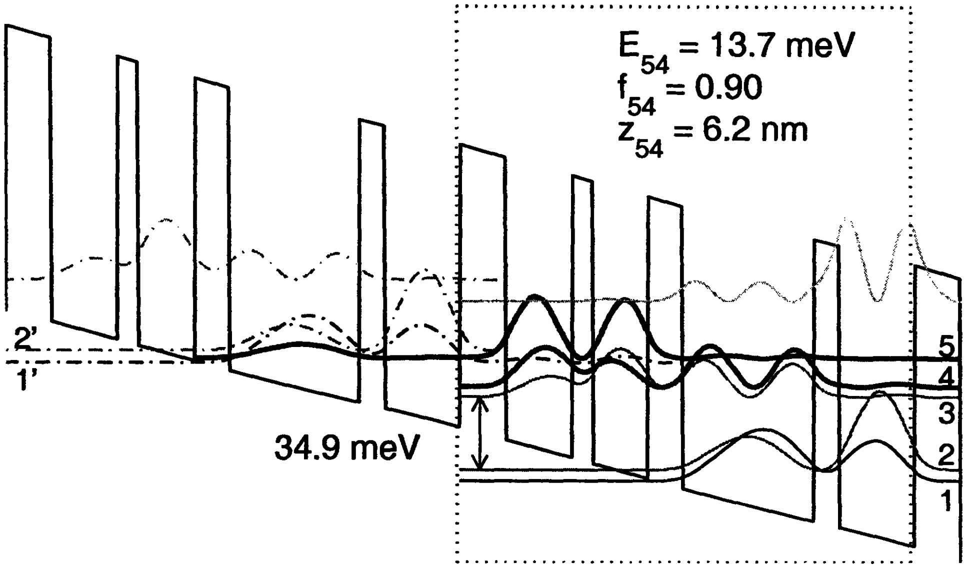

Figure 5-8: *Band structure of FL177C-M5. The device consists of GaAs/*Al$_{0.15}$Ga$_{0.85}$As *layers with thicknesses (nm)* **5.5**/*7.9*/**2.5**/*6.5*/**4.1**/*15.5*/**3.0**/*9.0 (barriers in boldface, wells in plain text) and a doping concentration* n*=1.9*$\times 10^{16}$ *cm*$^{-3}$ *in the 15.5-nm wide well, resulting in a sheet density of 3*$\times 10^{10}$ *cm*$^{-2}$ *per module.*

states n=1,2 is slightly larger than $\hbar\omega_{\mathrm{LO}}$, which ensures that the lower radiative state is depopulated via subpicosecond LO-phonon scattering. Table 5.2 and Fig. 5-9 show a comparison of some key simulation results with and without coherent transport (T_2=0.33 ps, to reproduce a FWHM linewidth of 6 meV together with relaxation scattering $\tau_4 \approx 0.5$ ps). From Fig. 5-9(a), we can see that the semi-classical results exhibit a large parasitic current at biases around 40 mV/module, corresponding to the $1' \rightarrow 3$ (with $\Delta_{1'3}$=0.45 meV) transition. The overestimation of the current density in this parasitic channel is due to the use of extended basis states in the calculation of the scattering rates. The density matrix approach largely eliminates this problem, but there is still a noticeable hump in the $I - V$ while there is none in the measurements, although the presence of the parasitic channel is still evident in differential conductance measurements. [57] This simulation result corresponds well with the simple NEGF calculations done for this same structure in Ref. [58] and reproduced here, which is a further indication that the used DM approach adequately models the coherence effects in the electron transport. The full NEGF simulation has a slightly reduced parasitic current density, but still overestimates J for all biases below injection anticrossing. For the calculations involving coherent transport, the predicted peak current density (700 A/cm^{-2}) and gain are achieved when the upper radiative level is lined up between the $2'$-5 and $1'$-5 anticrossings, and not at the $1'$-5 anticrossing as for the semi-classical simulation. Due to the finite linewidth of the levels ($>$ 4 meV for all simulations), which is comparable to $E_{2'1'} \approx 5$ meV, their similar anticrossing gaps ($\Delta_{1'5} \approx \Delta_{2'5} \sim 1.8$ meV) and the similar population densities of $1'$ and $2'$, injection is most efficient when both $1'$ and $2'$ contribute. Note that for small linewidths, injection from $2'$ and $1'$ into 5 is more centered around their respective anticrossing bias, which may result in a current peak at the $2' - 5$ anticrossing, followed by a region of negative differential resistance (NDR). In experiments, this early NDR would prevent the device from reaching its designed level alignment and degrade its performance.

The predicted peak population inversion ΔN_{54} and gain g are lower for the DM simulations than for the semi-classical calculations, ΔN_{54}=4.1$\times$10^9 cm^{-2} (g=32 cm^{-1} for Δf=6 meV) and ΔN_{54}=6.4$\times$10^9 cm^{-2} (g=52 cm^{-1}) respectively. This is largely due to a decreased injection efficiency and selectivity, as explained earlier in the discussion about the density matrix formalism. On the other hand, the NEGF simulations predict a narrower linewidth (Δf=2.5 meV) and a higher peak gain of about

Table 5.2: *Calculated subband energy, temperature, and population density of FL177C-M5 at the peak current density predicted by the DM simulation (V_{bias}=59 mV/module), for the semi-classical, density matrix and full NEGF simulations. The lattice temperature is 25 K.*

n	E (meV)	T_{el} (K)			Pop. (10^{10} cm^{-2})		
		semi-cl.	DM	Full NEGF	semi-cl.	DM	Full NEGF
1	0	114	143	111	1.14	1.27	0.91
2	5.4	127	138	102	0.74	0.93	0.76
3	41.3	152	174	109	0.13	0.10	0.17
4	46.3	154	174	97	0.19	0.11	0.26
5	60.4	115	108	104	0.73	0.52	0.84

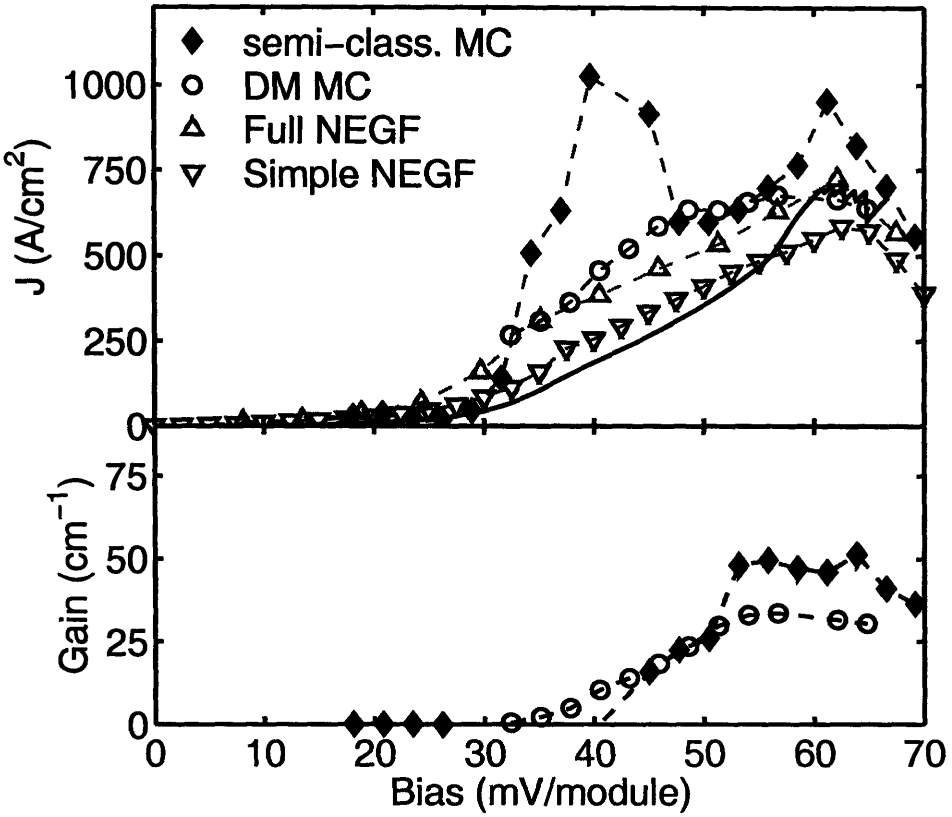

Figure 5-9: *(a) Current density for FL177C-M5 for a range of biases. The density matrix results were obtained for a phenomenological pure dephasing time* T_2*=0.33 ps. (b) Calculated gain for the MC simulations for a spontaneous emission linewidth of 6 meV.*

90 cm^{-1} for the full NEGF calculation (Fig. 5-10 and 105 cm^{-1} for the simple simulation (Fig. 5-11), which corresponds with 37 and 44 cm^{-1} respectively if the linewidth is adjusted to the experimentally observed Δf=6 meV. The discrepancy between the calculated and observed linewidths can be due to the underestimation of interface roughness scattering, or to inhomogeneous broadening which is not accounted for in the simulations.

The spatial and energetic distribution of electrons in the structure at injection anticrossing is shown in Fig. 5-13. The presence of a population inversion and the associated gain can easily be seen in the active region, where the population density of the upper radiative state is much higher than in the lower radiative state and resonant LO-phonon depopulated state. Also note that the populations of injector/collector

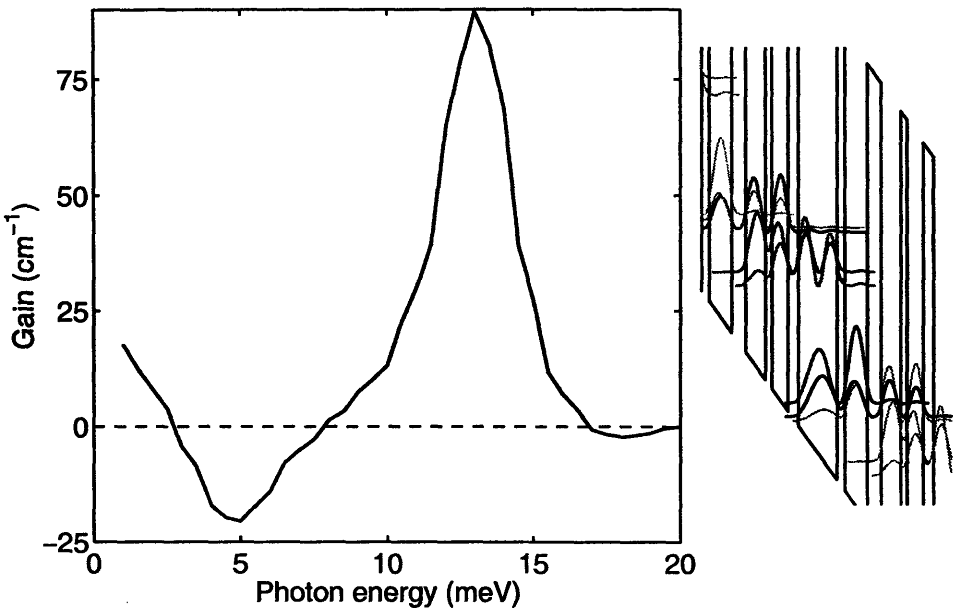

Figure 5-10: *Calculated gain spectrum for FL177C-M5 at injection anticrossing calculated with the full NEGF simulation. Also indicated are the diagonalized, extended wavefunctions at this bias.*

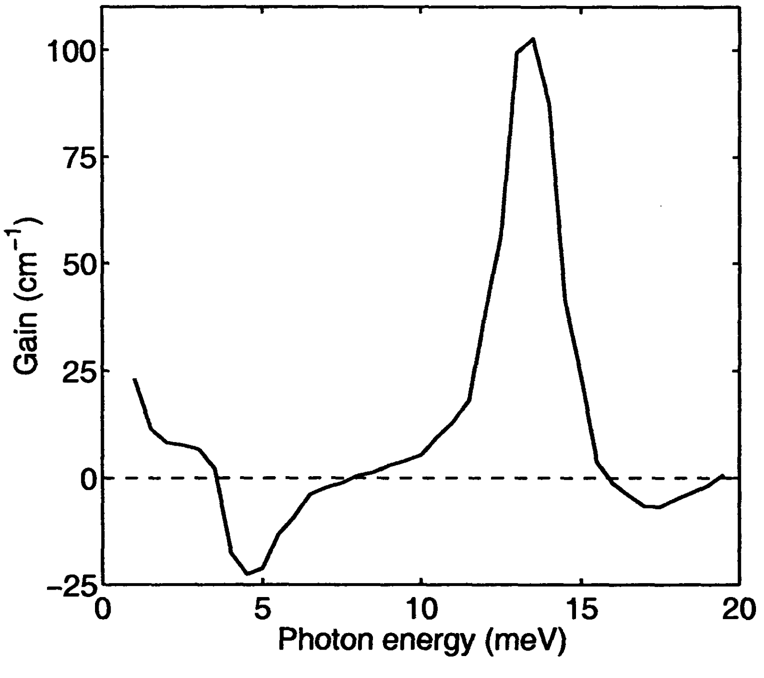

Figure 5-11: *Calculated gain spectrum for FL177C-M5 at injection anticrossing calculated with the simple NEGF simulation.*

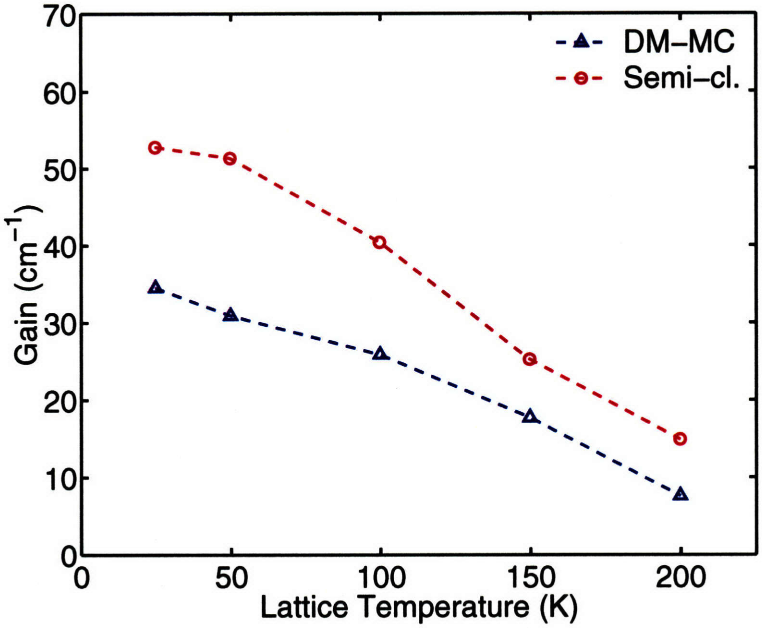

Figure 5-12: *Maximum gain of FL177C-M5 as a function of lattice temperature for the semiclassical and DM-MC simulations, assuming a spontaneous emission linewidth of 6 meV. In experiments, the lattice temperature is somewhat higher than the heat-sink temperature, and the devices lase up to ~ 130 K heat-sink temperature.*

states n=1 and n=2 cannot be distinguished in the figure. As was the case for T65, the population density in the injector well is highest, which means carriers are piling up behind the injector barrier and electron injection into the active region is affected by dephasing.

To investigate the high-temperature performance of the QCL, MC simulations for lattice temperatures up to 200 K were performed. As can be seen from Fig. 5-12, both simulations predict a steady decrease of the gain to g=8 cm^{-1} (DM) and 14 cm^{-1} (semi-cl.) at 200 K. This decline is mainly due to the increased LO-phonon mediated depopulation of level 5. In experiments, CW lasing was observed in very small devices (100×100 μm^2) immersed in liquid nitrogen, which roughly corresponds to a facet mirror loss of 40 cm^{-1}, for a facet reflectivity of ~ 60%. [59] This experimentally inferred gain agrees reasonably well with the predicted gain of ~ 30 cm^{-1} (DM) and 44 cm^{-1} (semi-cl.) at 77 K. The underestimation of gain and peak current density in the DM simulation indicates that the used pure dephasing time T_2 = 0.33 ps may be

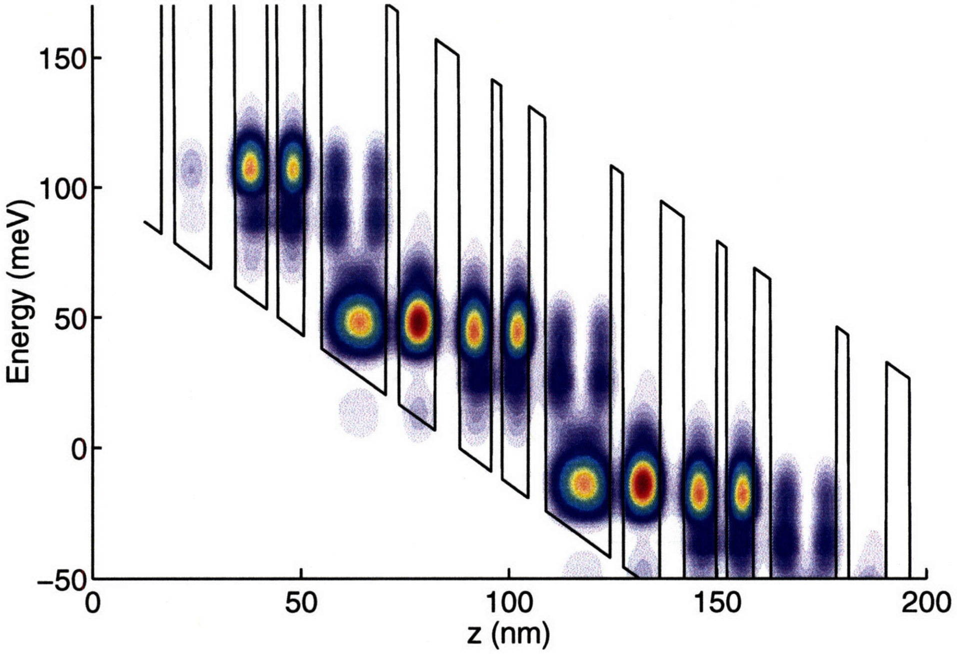

Figure 5-13: *Electron distribution in FL177C-M5 at injection anticrossing, assuming diagonalized wavefunctions. Indicated are the electron density in real space (horizontal axis) and energy space (vertical axis) for both subbands, using a color coding scheme ranging from red (high density) to blue (low density).*

too short.

Even though the difference between the peak gain and the current densities calculated from the semi-classical simulation and the simulations including coherent transport is only quantitative, there is an important qualitative difference in the current-voltage ($I-V$) characteristics. As can be seen in Fig. 5-9, the semi-classical current density at the $1'$-3 parasitic channel is higher than that at the designed bias of the $1'$-5 anticrossing. Experimentally, this would have resulted in a negative differential resistance (NDR) above the bias at the $1'$-3 anticrossing, making the desired $1'$-5 energy alignment inaccessible. Fortunately, dephasing reduces the current density at this parasitic channel much more than at the designed bias due to the smaller anticrossing gap $\Delta_{1'3} \approx 0.5$ meV, compared to $\Delta_{1'5} \approx 1.8$ meV. Consequently, we were able to bias the device at the $1'$-5 energy alignment and achieve lasing. However, for lower-frequency QCLs at 2.1 THz, the biases of $1'$-3 and $1'$-5 are closer, resulting in much more similar current densities at those respective bias conditions and a much smaller range of current densities in which the devices lase. [60] In this important aspect, the semi-classical MC calculation is useless in predicting the relative current densities at the $1'$-3 and $1'$-5 biases, and whether a NDR will occur below the design bias. With the much smoother $I - V$ that is closer to the experimental results, the DM-MC and NEGF simulations could help us in designing suitable structures with a reduced $1'$-3 parasitic current density for lower-frequency lasers.

5.3 OWI185-M1

The third device is a QCL which lased at 1.86 THz [7], and whose band diagram and wavefunctions are reproduced in Fig. 5-14. In order to prevent reabsorption of THz radiation due to intersubband transitions and to enhance the injection selectivity, this device was designed with a one-well injector region and a single injector subband. At resonance, electrons are injected into the upper level n=4 and make a radiative transition down to n=3, which is in resonance with level 2. The latter subband is depopulated by resonant LO-phonon scattering into n=1, from where they are injected into the next module. The small anticrossing gap between the injector state and upper radiative state ($\Delta_{1'4}$ ≈1 meV) combined with dephasing scattering makes transport through the injection barrier incoherent. The energy separation between the upper radiative level n=4 and the "parasitic" level n=5 is E_{54} ≈14 meV, which is much larger than its anticrossing energy with the injector state $\Delta_{1'5} = 1.5$ meV, and no electrons are injected directly from n=1' into n=5. This means that n=5 will play only a very limited role in the electron transport, but could get populated with hot electrons from n=4, thus reducing gain at high temperatures.

It should be pointed out that there is a substantial uncertainty about the reported doping density $n^{\mathrm{OWI185-M1}}$=2.25×10^{10} cm^{-2} of this device. Due to a problem during the growth, the doping density was not set correctly and still remains in doubt. However, by comparing the peak current density of this device with that of another, very simar device (OWI185), we can get an idea of the possible error in $n^{\mathrm{OWI185-M1}}$. Assuming a similar fraction of the population density is residing in the injection subband

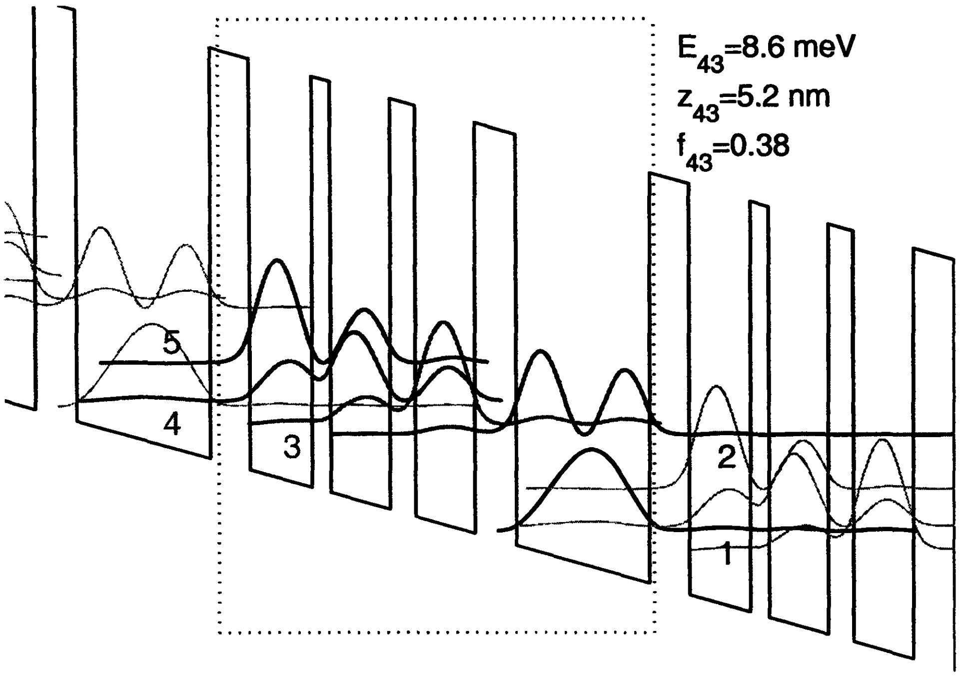

Figure 5-14: *Band structure of OWI185-M1 at the injection anticrossing (45.6 mV/module), obtained from the NEGF calculations, with the four-well module outlined in a box. The device consists of GaAs/*$\mathrm{Al_{0.15}Ga_{0.85}As}$ *layers with thicknesses (nm)* ***4.9****/7.8/****2.3****/7.6/****3.2****/7.6/****5.2****/16.8 (barriers in boldface, wells in plain text) and the 5.2-nm wide barrier is delta-doped in the center at 2.25×*10^{10} *cm*$^{-2}$*.*

and dephasing scattering dominates the transport bottleneck through the injection barrier, the peak current densities for both devices are determined by Eq. 2.81 at resonance and their ratio can be written as:

$$\frac{J_{\mathrm{peak}}^{\mathrm{OWI185-M1}}}{J_{\mathrm{peak}}^{\mathrm{OWI185}}} = \frac{n^{\mathrm{OWI185-M1}}}{n^{\mathrm{OWI185}}}\left|\frac{\Delta_{1'4}^{\mathrm{OWI185-M1}}}{\Delta_{1'4}^{\mathrm{OWI185-M1}}}\right|^2 . \tag{5.1}$$

The measured current densities were very close, $J_{\mathrm{peak}}^{\mathrm{OWI185-M1}}$=186 A/cm^2 and $J_{\mathrm{peak}}^{\mathrm{OWI185}}$=170 A/cm^2, while the injection anticrossings were $\Delta_{1'4}^{\mathrm{OWI185-M1}}$ =1 meV and $\Delta_{1'4}^{\mathrm{OWI185}}$=0.76 meV, which are both small enough to justify the assumption that transport is determined by sequential tunneling. For a doping density of $n^{\mathrm{OWI185-M1}}$= 2.7×10^{10} we therefore expect a current ratio of 1.45, while the measured ratio is 1.09. This suggests that the actual doping density $n^{\mathrm{OWI185-M1}}$ could be as low as 1.7×10^{10}, which has to be taken into account when comparing the simulation results to the measurements.

The calculated current densities for the semi-classical MC, density matrix MC

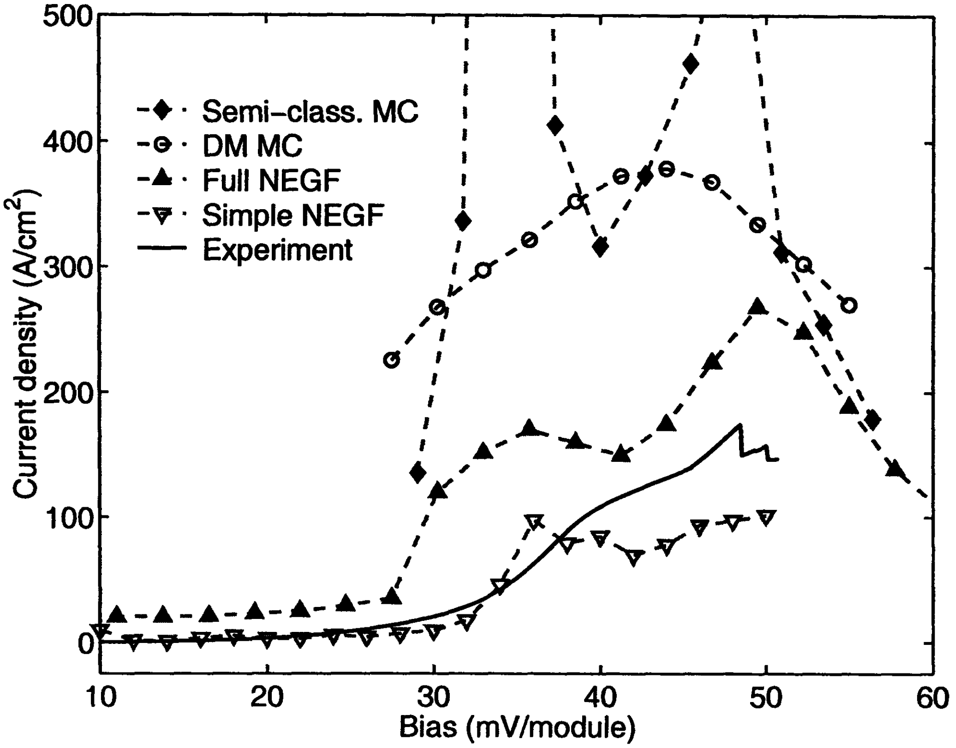

Figure 5-15: *Current density for OWI185-M1 for a range of biases (total electron density* n*=2.25*$\times 10^{10}$ *cm*$^{-2}$*. The density matrix results were obtained for a phenomenological pure dephasing time* T_2*=0.33 ps.*

(T_2=0.33 ps to reproduce a FWHM linewidth of 6 meV with relaxation scattering $\tau_4 \approx 0.5$ ps) and both NEGF simulations are shown in Fig. 5-15. The Monte Carlo simulations both severely overestimate the current density for all biases, while the NEGF calculations yield a closer match. All simulated current densities exhibit a bump around 35 mV/module, especially the semi-classical simulation which exhibits a very large peak $J \approx 2500$ A/cm^{-2}. In this bias range, the injector level lines up with n=2 ($\Delta_{1'2}$ =0.3 meV) and n=3 ($\Delta_{1'3}$=0.35 meV), creating a parasitic current channel. The electrons injected into n=2,3 are depopulated by resonant LO-phonon scattering into the next injector level n=1, which results in a large parasitic current despite the small anticrossing gaps. Similar to FL177C-M5, the overestimation of the current density in the semi-classical simulation is caused by the use of extended wavefunctions. The peak current density for the DM simulation occurs at a bias slightly past the parasitic bias, when the injector energy is in between n=3 and n=4

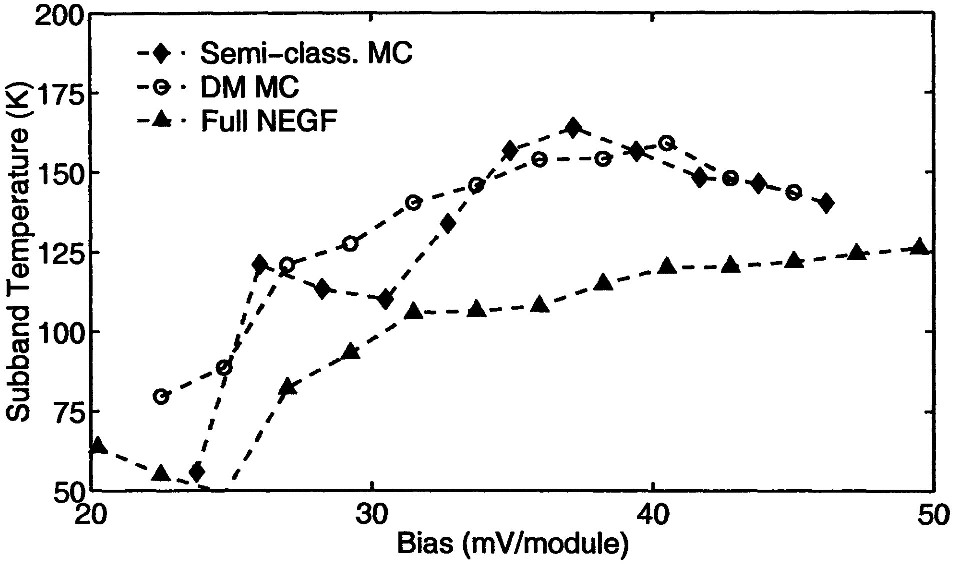

Figure 5-16: *Subband temperature of the upper radiative level $n=4$ for a range of biases for the semiclassical, density matrix and full NEGF simulations. The lattice temperature was 25 K.*

and injects electrons into both levels. There is however no evident explanation for the overestimation of J for the DM calculation. The difference between the simple and full NEGF simulations is mainly due to the underestimation of LO-phonon scattering out of n=3 and n=2 in the simple simulation; furthermore, the absence of e-e scattering in the simple simulation between the closely spaced subbands n=2 and n=3 further reduces the current density. However, rather than only a quantitative there is also a

Table 5.3: *Calculated subband energy, temperature, and population density of OWI185-M1 at the peak current density predicted by the full NEGF simulation (V_{bias}=45.6 mV/module), for the semi-classical, density matrix and full NEGF simulations. The lattice temperature is 25 K.*

n	E (meV)	T_{el} (K)			Pop. (10^{10} cm^{-2})		
		semi-cl.	DM	Full NEGF	semi-cl.	DM	Full NEGF
1	0	139	128	105	0.94	1.52	1.05
2	35.6	158	145	117	0.05	0.10	0.06
3	39.9	156	154	119	0.16	0.13	0.18
4	48.1	130	125	144	0.88	0.30	0.86
5	62.1	95	85	95	0.22	0.20	0.10

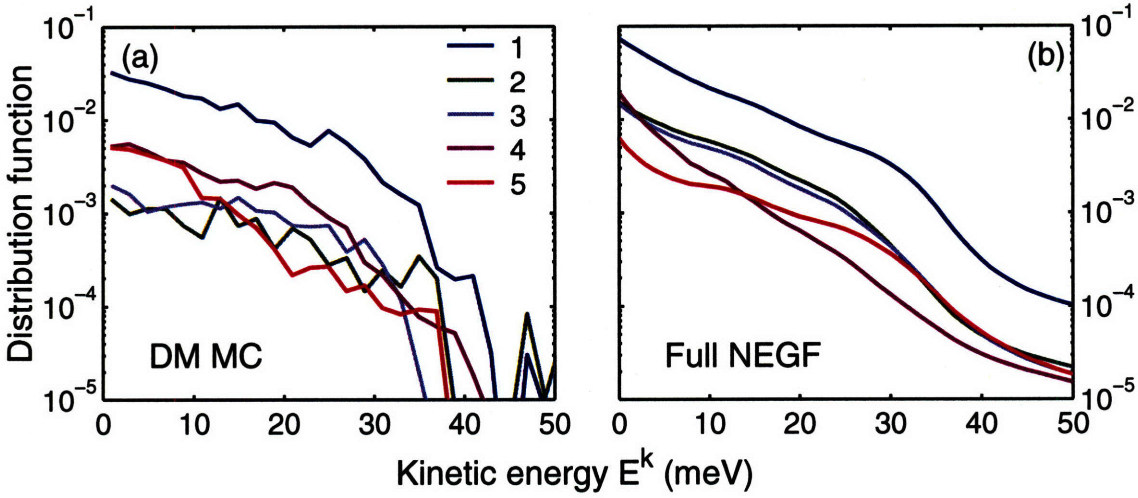

Figure 5-17: *Electron distribution functions for OWI185-M1 at injection anticrossing for (a) the DM MC simulation and (b) the full NEGF simulation (Wannier functions). Due to the different set of basis wavefunctions a direct comparison between (a) and (b) is not possible. The lattice temperature was 25 K for both calculations.*

qualitative difference between both NEGF results. Since the calculated parasitic and design peak current densities are very similar for the simple NEGF, the predicted occurrence of an NDR region at biases higher than the parasitic bias would make the 1'-4 anticrossing inaccessible and significantly degrade the device's performance. On the other hand, for the full NEGF simulation, the parasitic J_{peak}=175 A/cm^{-2} is much lower than the predicted design J_{peak}=270 A/cm^{-2}, and an NDR would have a very limited impact.

Table 5.3 summarizes the calculated population density and subband temperature for the MC and full NEGF calculations at injection anticrossing. The subband populations for the DM MC simulation refer to localized wavefunctions, whereas the semi-classical and NEGF results refer to extended wavefunctions. Only a modest number of electrons, 5-10% of the total density, are populating the parasitic level n=5. As can also be seen in Fig. 5-16, the NEGF simulation predicts lower subband temperatures than the MC simulations. This is due to the incorporation of energy spectra in the NEGF formalism, which increases the number of states available for e-LO scattering and hence energy relaxation and in turn results in a lower electron temperature. The electron distributions of most subbands are thermal for $E^k < 30$

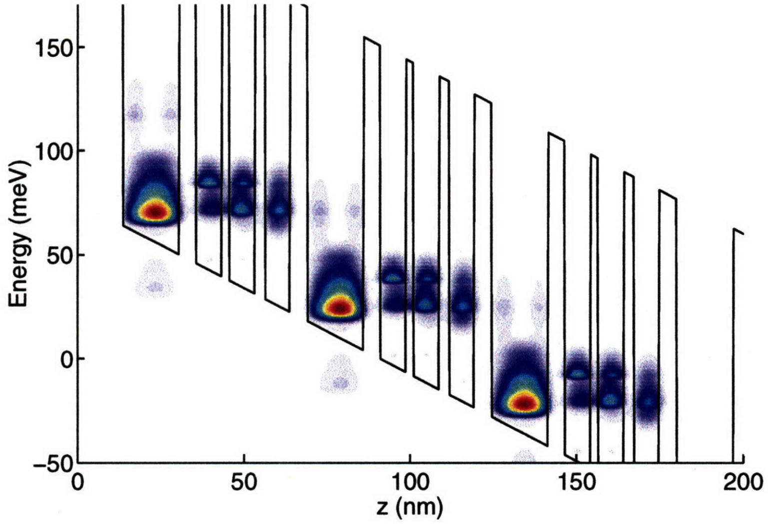

Figure 5-18: *Electron distribution in OWI185-M1 at injection anticrossing, assuming diagonalized wavefunctions. Indicated are the electron density in real space (horizontal axis) and energy space (vertical axis) for both subbands, using a color coding scheme ranging from red (high density) to blue (low density).*

meV (Fig. 5-17). The onset of fast intrasubband LO-phonon scattering causes a sharp decrease in the distribution functions above E^k=36 meV. Note that the shown NEGF distribution refer to the Wannier functions and not the extended wavefunctions, and therefore a direct comparison between the DM and NEGF results is only possible for n=1 since this subband is nearly identical in both cases. Figure 5-18 shows a graphical representation of the spatial and energetic distribution of electrons in the device, based on the full NEGF simulation. It can clearly be seen that a majority of the carriers is stuck in the injector and that the injector barrier is a major impediment for transport. Also visible are the electron distributions of the upper radiative level n=4 and the parasitic level. On the other hand, the resonant-LO-phonon depopulated level n=2 is nearly empty and is only visible as two faint stains in the wide well.

Similar to FL177C-M5, the DM MC simulation predicts a much lower peak population inversion ($\Delta N_{43} = 1.7 \times 10^9$ cm^{-2} and gain (g=7 cm^{-1} for Δf=6 meV) than the semi-classical calculation ($\Delta N_{43} = 6 \times 10^9$ cm^{-2} and g=28 cm^{-1}) (Fig. 5-19). The gain spectrum obtained from the full NEGF simulation at injection anticrossing gives a more complete picture. The 4-3 population inversion results in a gain peak of 27

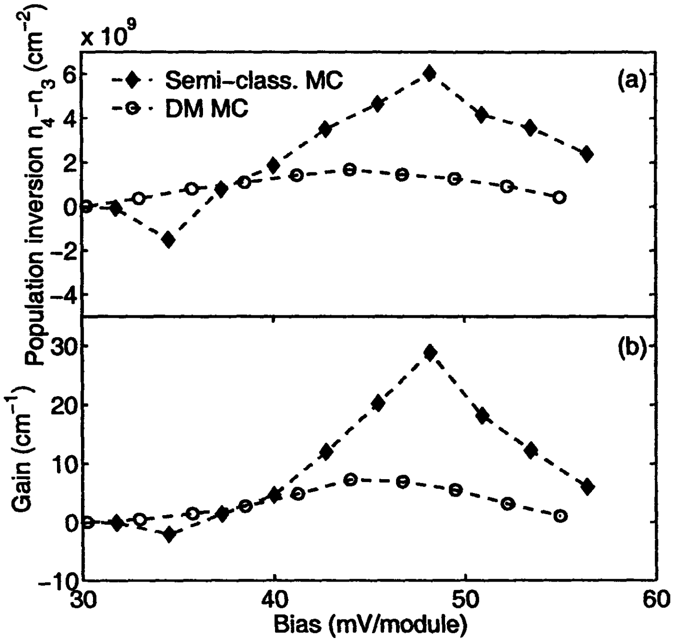

Figure 5-19: *(a) Calculated population inversion $N_4 - N_3$ for the MC simulations of OWI185-M1. (b) Predicted gain for OWI185-M1 for a spontaneous emission linewidth of 6 meV.*

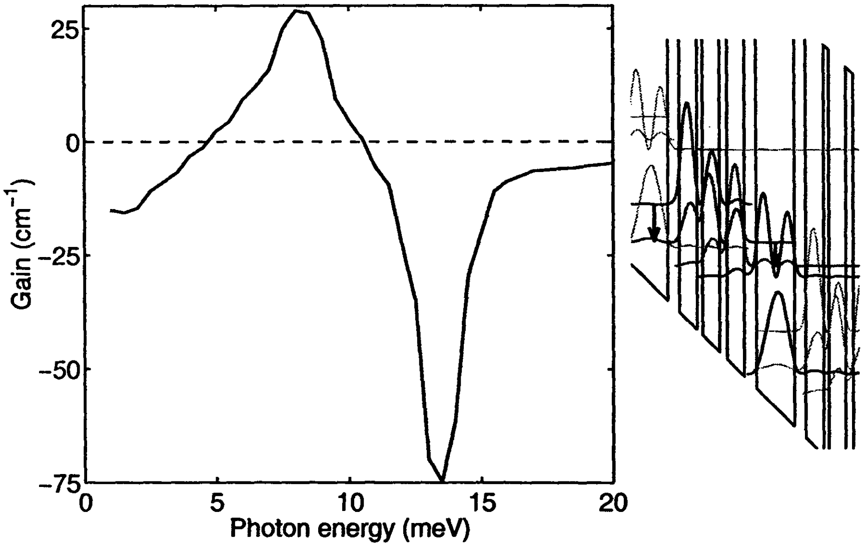

Figure 5-20: *Calculated gain spectrum for OWI185-M1 at injection anticrossing calculated with the full NEGF simulation. Also indicated are the diagonalized, extended wavefunctions at this bias.*

n	E (meV)	T_{el} (K)		Pop. (10^{10} cm^{-2})	
		without imp.	with imp.	without imp.	with imp.
1	0	102	121	0.77	0.98
2	6.5	111	139	1.29	0.95
3	44.0	122	145	0.10	0.12
4	49.3	133	160	0.08	0.12
5	63.3	96	117	0.64	0.70

Table 5.4: *Calculated subband energy, temperature, and population density of the 3.4-THz laser at injection anticrossing (T_{latt}=25 K) with and without impurity scattering.*

cm^{-1} at $\hbar\omega$=8 meV with a linewidth of 2.5 meV, which is equivalent to a peak gain of 11 cm^{-1} for Δf=6 meV. The narrow absorption peak (Δf=1.5 meV) at $\hbar\omega$=14 meV is due to the transition between the more densely populated upper radiative state and the parasitic level n=5. Experiments and Drude model calculations suggest that the combined waveguide and mirror losses of the investigated device add up to approximately 5-10 cm^{-1} [7], which is close to the calculated DM and NEGF gain.

5.4 Impurity scattering

To illustrate the importance of impurity scattering in electron transport in THz QCLs, Monte Carlo (MC) simulations [12] were used to investigate several different QC designs. [61] Here we focus on two representative examples, for which experimental results were published. All simulations include e-ph (acoustic and LO-phonon) and e-e scattering. Calculations were performed with and without e-imp scattering. A non-equilibrium, multi-subband screening model [62] was used for e-imp and e-e interactions. No phenomenological parameters were introduced. When comparing the calculation results with the measurements, the current density J provides a good reference point. Only when the simulation produces current densities consistent with experiments can we have confidence in other calculated results, such as subband populations and gain. All scattering times are net scattering rates and include the effect of backfilling. [12]

The first investigated device is FL175C, a 3.4-THz QCL [11] very similar to

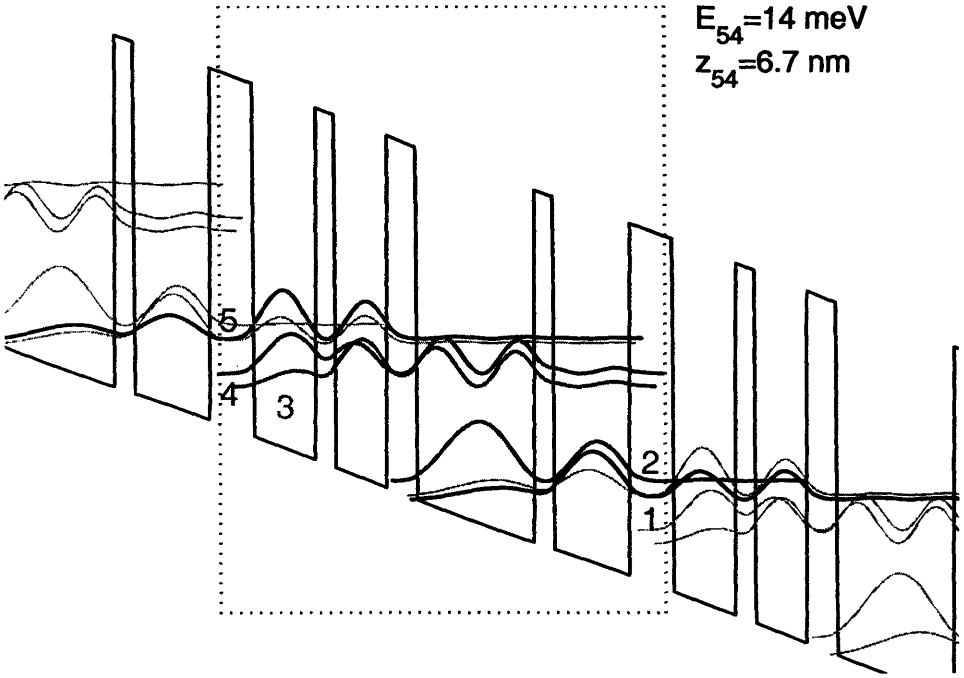

Figure 5-21: *Band structure for FL175C. The device consists of GaAs/*$\mathrm{Al}_{0.15}\mathrm{Ga}_{0.85}\mathrm{As}$ *layers with thicknesses (nm)* **5.4**/*7.8*/**2.4**/*6.4*/**3.8**/*14.8*/**2.4**/*9.4* *(barriers in boldface, wells in plain text) and is doped to* $n=1.9\times10^{16}$ cm^{-3} *in the 14.8 nm wide well, resulting in a sheet density of* 2.8×10^{10} cm^{-2} *per module.*

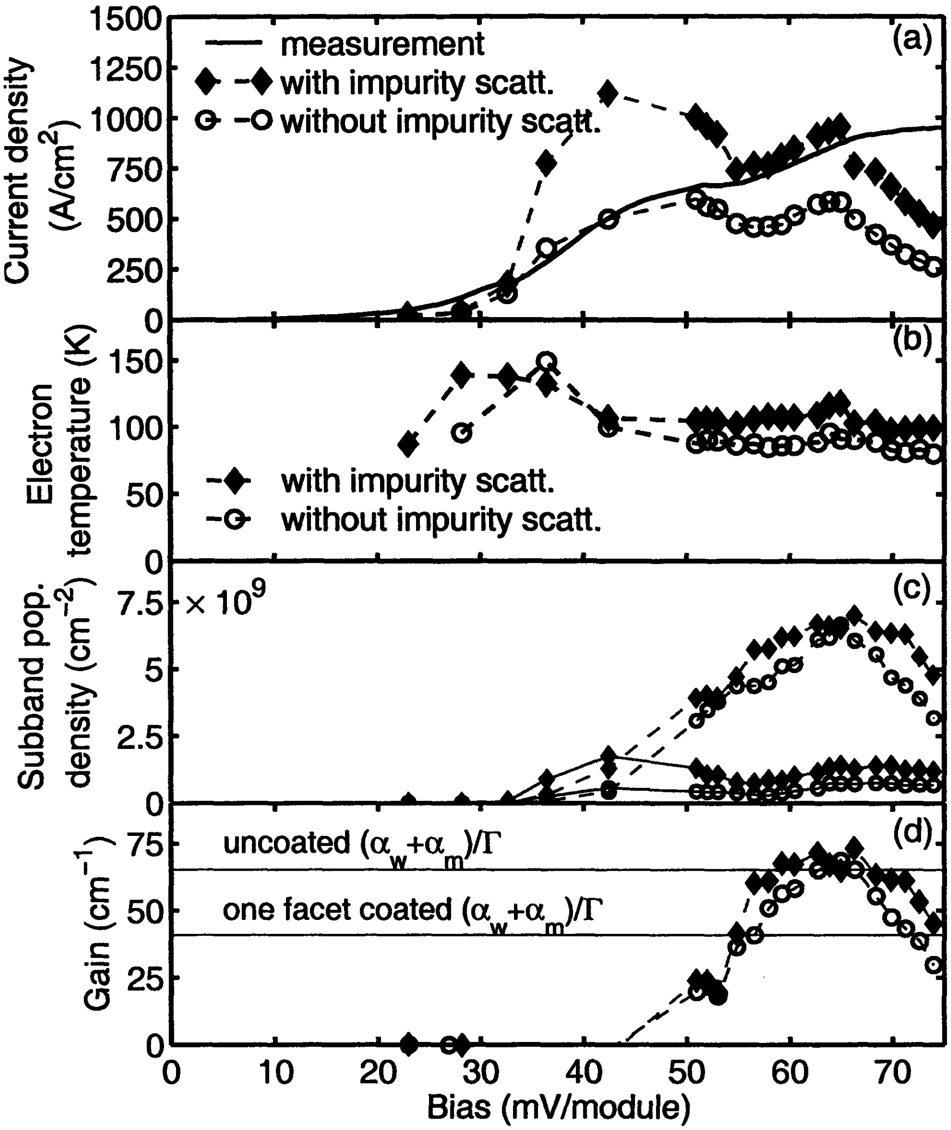

Figure 5-22: *Key results of the MC simulation of FL175C for T_{latt}=25 K, calculated with and without e-imp scattering (represented by diamonds and circles, respectively). (a) J for a range of biases. The measured current density (solid line) was adjusted to account for a parasitic series resistance of 2 Ω. The large parasitic current peak of ~2000 A/cm^2 at ~45 mV/module was omitted from the calculation results. (b) T_{el} for n = 5, the upper radiative level. (c) The population density in n=4 (solid line) and n=5 (dashed line). (d) Material gain for a 1180×150 μm^2 ridge structure. The two horizontal lines represent the total cavity losses with uncoated facets and with one facet fully reflecting.*

FL177C-M5, which was discussed earlier.(Fig.5-21) Table 5.4 and Fig. 5-22 show a comparison of the key MC results with and without e-imp scattering, for a lattice temperature T_{latt}=25 K. Note that the inclusion of e-imp scattering results in an increase in electron temperature. Also, the calculated J at injection anticrossing ($\sim$ 65 mV/module) increases from 580 A/cm^{-2} to 950 A/cm^{-2}. The simulation result closely tracks the measured current density for biases larger than 55 mV/module, in the range where no large parasitic current channels [12] are present. The increase in J is due to an enhanced scattering rate in the injector region (between n=2' and n=1'), and between n=1' and the upper radiative state n=5. τ_{21} decreases from 14 ps to 5 ps due to τ_{21}^{imp}=7 ps, eliminating the population inversion between n=2' and n=1', in contrast to the result in Ref. [1]. In spite of the decrease in $\tau_{1'5}$=4.5 ps (from 6 ps without e-imp scattering), the population density of n=5 remains almost unchanged because of the increased thermally activated LO-phonon scattering ($\tau_5^{LO} \approx$ 2.8 ps from 4.1 ps). Note that, although e-imp scattering between the radiative levels (τ_{54}^{imp}=23 ps) is faster than e-e scattering (τ_{54}^{e-e}=49 ps), its (direct) contribution to the depopulation of n=5 is negligible. τ_4 is dominated by resonant LO-phonon scattering, and is little affected by e-imp scattering. However, the rise in J results in a slight population increase in n=4. Using measured values [63] for the refractive index n=3.8 and the linewidth $\Delta\nu$=1 THz, and the calculated population inversion ΔN_{54}=5.8$\times$10^{10} cm^{-2}, a peak gain of 73 cm^{-1} is found (68 cm^{-1} without e-imp scattering). This value slightly exceeds the upper limit of the gain range inferred from experiments, which is most likely due to the neglect of wavefunction localization caused by incoherent transport through the injector barrier (anticrossing $\Delta_{1'5} \approx$ 1.8 meV).

The second QC device is the simple double-quantum-well structure T65, [56] which was discussed earlier. In experiments, the current characteristics of this device were nearly independent of temperature for T_{latt}=5 – 77 K, while the electroluminescence dropped sharply. This is a strong indication that transport through the thick injector barrier is limited by incoherent tunneling between the injector state and the upper radiative level (anticrossing $\Delta_{1'3} \approx$ 0.8 meV). In such a case, the use of extended wavefunctions for transport analysis would be expected to overestimate J. Table 5.1

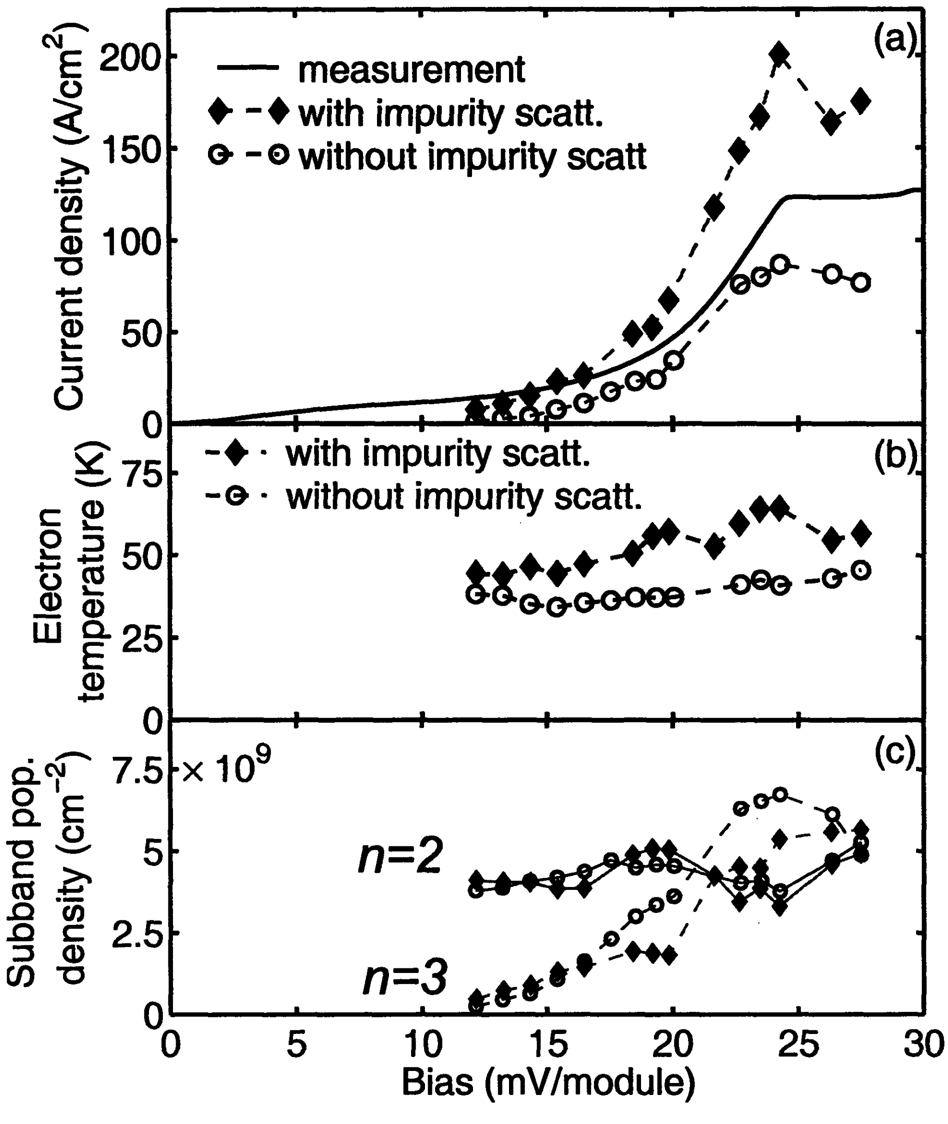

Figure 5-23: *Key results of the MC simulation of the two-well structure for* T_{latt}*=25 K, calculated with and without e-imp scattering (represented by diamonds and circles, respectively). (a) J for a range of biases. The measured current density (solid line) was adjusted to take into account a parasitic series resistance of 1.5* Ω*. (b)* T_{el} *for* $n = 3$*, the upper radiative level. (c) The population density in n=2 (solid line) and n=3 (dashed line).*

and Fig. 5-23 present the main calculation results with and without e-imp scattering, for T_{latt}=25 K. Fig. 5-23(a) shows that the inclusion of e-imp scattering is necessary to obtain the expected overestimation of the peak current density. As with the 3.4-THz laser, the increase in current can be explained by the enhancement of transport through the collector state due to e-imp scattering, combined with a decline in τ_3 caused largely by thermally activated LO-phonon scattering ($\tau_3^{LO} \approx 13$ ps, compared to 35 ps without e-imp scattering). The calculated lifetime of n=2 is reduced from $\tau_2 \approx \tau_2^{e-e}$=13 ps to 4 ps by the inclusion of τ_2^{imp}=6 ps. The poor agreement between calculation and experiment for this device is likely due to the small anticrossing gap $\Delta_{1'3} \approx 0.8$ meV.

Chapter 6

Conclusion

In this thesis, I have described different approaches, both semi-classical and quantum-mechanical, to model electron transport in quantum cascade lasers and MQW devices in general. Over the course of my PhD, I have developed simulation tools implementing these different approaches, and compared the results with experimental data to determine its accuracy and to investigate how to improve the models. Starting from a simple semi-classical model, which treated electron transport with essentially a set of rate equations describing only e-e and e-phonon scattering, we gradually added more scattering mechanisms (e-imp, e-interface roughness) and resonant tunneling by means of the density matrix formalism. A variety of quantum cascade structures were simulated with this code, and the results were compared with experiments. Encouraged by Andreas Wacker's results with the NEGF approach, we made a major effort to develop a new simulation based on the papers written by him and his co-workers to obtain an even more comprehensive picture of electron transport.

Some of the results and conclusions we obtained from these calculations are summarized in what follows.

Our first semi-classical Monte Carlo calculations of a working THz laser (FL175C), including only e-e and e-phonon scattering, yielded current densities that were significantly lower than observed, as well as a gain that was higher than inferred from experiments. However, due to the spatially extended wavefunctions used in the semi-classical picture and the lack of dephasing, the calculated current density should always exceed the experimental values. This suggested the presence of a major scattering mechanism that was yet unaccounted for in this model. The addition of electron-impurity scattering proved to be the solution, resulting in the expected overestimation of the current density. We were able to show that the importance of e-imp scattering in electron transport in THz QC structures rivals or even exceeds that of e-e scattering, and that its inclusion in the calculations is crucial when modeling the intersubband transport dynamics in these devices.

As a second step we have shown that the inclusion of a model for coherent transport and dephasing is essential to describe the transport dynamics of intersubband transport in THz QCLs. We implemented this quantum transport with two different formalisms.

The first approach consisted of adapting the semi-classical simulation to include a density matrix description of transport. The density matrix model, together with the choice of localized basis states, allows for an intuitive treatment of transport between weakly coupled levels and the incorporation of the effects of sequential tunneling into a Monte Carlo simulation. Scattering events, including e-phonon, e-e and e-imp scattering, are treated semi-classically but contribute to dephasing scattering. In addition, a phenomenological "pure dephasing rate" was introduced to take into account dephasing caused by interface roughness scattering. We have used the semi-classical and the density matrix MC simulation to compare calculated current densities and gain with experimental measurements. The inclusion of coherent transport showed marked improvement over the semi-classical model. It largely eliminated the overestimation of the peak current density and parasitic current channels, and correctly predicted the absence of a population inversion where the semi-classical model predicted a large gain. However, more remains to be done. The use of a single, phenomenological pure dephasing time to describe the interaction between all subbands does not take into account the substantial differences in elastic intrasubband (impurity and interface roughness) scattering for different levels in a module. The different models used to describe these scattering mechanisms and the large scattering rates these models predict for small-angle scattering, make it difficult to dependably estimate their contribution to dephasing. More detailed calculations incorporating this subband-dependant dephasing can provide a model that uses no phenomenological parameters, and will yield a more accurate description of the electron transport. This more comprehensive simulation can be a valuable tool for designing and analyzing QCLs.

As an alternative to the density matrix approach, a simulation based on the non-equilibrium Green's function formalism was developed. In contrast to the previously mentioned semi-classical and DM models, the NEGF formalism does not provide a clear and intuitive picture of the physics underlying electron transport, which makes it much harder to obtain a good understanding of a device's inner workings and could pose a barrier to its use as a design tool. However, the formalism naturally describes the energy spectrum of the states involved in transport, including important attributes such as linewidth and lineshape, which are lacking in the DM picture. Furthermore, scattering is handled quantum-mechanically and takes into account the coherent interaction between states, as opposed to the semi-classical approach used in the DM picture. The resulting physical model is comprehensive but opaque, and its implementation as a simulation is very computation intensive.

The NEGF formalism describes coherent electron transport including the effects of e-phonon, e-imp and e-interface roughness scattering. Two implementations of the NEGF formalism, with momentum-dependent and -independent scattering matrix elements respectively, were used to model existing QC structures. The simplified simulation with momentum-independent scattering is less accurate, but much less computation intensive than the complete calculation. For the complete NEGF simulation, a model for e-e scattering was added and verified. Current densities, population distributions and gain spectra were calculated and compared with similar results from the Monte Carlo simulations. In general, the NEGF simulations give a more accurate prediction of the experimental results than the MC results. The results for both NEGF simulations are similar enough that a good calculation strategy to investigate a new device can consist of a "rough", quick first pass with the simple simulation to assess its suitability as a QCL, followed up by a complete simulation at various bias voltages to investigate the transport in more detail.

A full, comprehensive study of the predictive power of all simulations (especially NEGF) could not be completed due to time restrictions. However, with a limited amount of effort, improvements can be made to both the MC and NEGF simulations that make them more comprehensive and allow for an easier access to some results of interest.

The addition of an improved model for interface roughness scattering to the MC and NEGF simulations could be highly beneficial for the accuracy of both models, and may eliminate the need for a phenomenological scattering time in the density matrix MC simulation. The current model tends to generate very high semi-classical scattering rates at shallow angles (approaching infinity at $\theta=0$), and is only used to describe intersubband scattering in the MC simulations while intrasubband interface roughness scattering is modeled through the phenomenological T_2. On the other hand, in the NEGF formalism interface roughness scattering contributes so little to the self-energies that it can be neglected without affecting the results. This apparent underestimation of interface roughness scattering is partially responsible for the underestimation of the spontaneous emission linewidth in the NEGF simulations. If these concerns can be addressed with an updated model for interface roughness scattering, the predictive power of the simulations could be significantly improved.

As the maximum operating temperature of QCLs continues to increase, modeling the devices at higher lattice temperatures and predicting their temperature performance will become increasingly important. Unfortunately, neither the MC nor the NEGF simulations include a working model of acoustic phonon scattering, which is very temperature sensitive and becomes a major transport mechanism in QCLs above a lattice temperature of 150-200 K. The energy dispersion of acoustic phonons makes this problem much harder to treat than LO-phonons. However, it is possible to approximate acoustic phonons as dispersionless "low-energy LO-phonons" with an energy that depends on the lattice temperature; the accuracy and importance of such an addition to the simulation needs to be investigated before any reliable high-temperature predictions can be made.

Due to time constraints in the development of the code, many important calculation results from the NEGF simulations are not readily available. Based on the description of the intersubband currents in section 4.7, it is straightforward to calculate scattering lifetimes for the different scattering mechanisms for any given intersubband transition as well as the contribution from coherent interaction. By diagonalizing the Hamiltonian, these scattering rates and currents can be calculated for states that fit well in an intuitive picture of electron transport, which would help the interpretation of the results immensely. Finally, some additional optimization of the memory management and algorithm routines can help reduce the computational overhead and allow for a faster and more accurate simulation result.

Appendix A

Appendix: Fortran Code

A.1 Monte Carlo Simulation

The MC simulation consists of one file of source code, named *eeisb_tb14.F*, and an input file *mcinput.dat*. The input file can be generated with the help of the matlab script *mclayersdm.m*, and contains information regarding the layer structure, doping, how the wavefunctions are localized, the lattice temperature and the applied bias. As an example, we show the first part of *mclayersdm.m* which creates the input file for a simulation of FL175C (the second part, in which the actual file is generated, is omitted):

```
scaling=0.92; %scaling factor to account for under/overgrowth

%monolayer thickness (in Angstrom)
d=2.825;

%number of modules
nmmod=1;

%thickness of boundary barriers in monolayers
tbarr=30;
tbarr=tbarr*d;

%module layer thicknesses in monolayers, starting with a well and ending in a barrier
l=[90 8 51 19]*scaling;
l=l*d;

%number of layers
nmreg=length(l);

%number of submodules
nmsmod=1;

%number of wells per submodule (!!! wells, not layers!!)
nmwells=[2];

% Al fraction in layers
xmol=0.15;

%doping regions
%  Doping is entered in matrix, dope(region,specification)
%  where region is an integer corresponding to a region number
%        dope(reg,1)=xmin (angstroms)
%        dope(reg,2)=xmax (angstroms)
%        dope(reg,3)=doping concentration (cm-3)
%        dope(reg,4)=Charge type (1.0 donors or -1.0 acceptors)
%  ndreg is the number of different uniformly doped regions
%  NOTE: For some reason, I use the interface between the first and
%   second region as the point z=0.
ndreg=1;
```

```
dope(1,1)=98*d*scaling;
dope(1,2)=dope(1,1)+51*d*scaling;
dope(1,3)=1.4e+16;
dope(1,4)=1.000;

%electric field in V/m
wfield=-540000.000;

% max number of levels per module
nsmxmd=4;

% lattice temperature (K)
tl=25;

% simulation duration (s)
tsimax=80e-12;

% 2D electron density (MKS)
econ=2.017e14*scaling;

% Conduction band offset percentage
cboffset=0.72;

% form factor jumper
jloadff=0;

% jumper to enable dynamic updating of subbands
jrecalc=0;

% jumper to enable wavefunction cropping
jcropwf=0;

%%%%%%%%%%%%write input deck%%%%%%%%%%%%

(continued)
```

Before compiling the source code, we need to make sure that enough memory is available for the proper execution of the program:

```
ulimit -s unlimited
```

The code can then be compiled using the Fortran compiler:

```
f77 -cpu:p7 -W -O3 -N113 -lg2c -llapack eeisb_tb14.F
```

The output file is *a.out*. In the following sections, I will briefly summarize the different subroutines in the Monte Carlo code.

absc

This subroutine is obsolete, but was left in because it initializes some variables. It can probably be thrown out entirely if these variables are initialized elsewhere.

acous

Calculation and rejection of the final state for acoustic phonon scattering, assuming a dispersionless acoustic phonon an energy equal to the lattice thermal energy. Input parameters are:

e: energy of the scattered electron

kx,ky: momentum of the scattered electron

n: initial subband

m: final subband

iv: band to which the particle belongs, always equal to 1 (conduction band) *part_wt*: particle weight, i.e. a weighting factor to determine how much of the density matrix describing the electron is scattered

aqee

Determination of the e-e form factors, saved in the variable *feei.* To speed things up, the algorithm first calculates another form factor that recurs in many calculations. Only the form factors that are larger than a certain threshold value are evaluated explicitly, the others are set to zero. As a final step, the subband indices involved in the non-zero form factors are recorded in the arrays *numinteree, isubinteree* and *numee* far easy access later. The form factors are written out to the file *formfactors.dat.*

aqnm

Determination of the form factors for e-imp and e-LO-phonon scattering. Output files are *ffim.dat* (e-imp) and *ffpop.dat* (e-LO-phonon scattering).

avg

Determines the average and variance of a variable x. Input/output variables:

x: vector array of the parameter to be averaged

n1,n: begin and end indices of x, indicating the part of x which is averaged

avrg,var: output variables for the average and variance

calcxnn

Calculates the fraction *frcar* (expressed as a number 0-1) and *xnn* (in percent) of the total number of carriers in a given subband.

defreg

Reads the input file *mcinput.dat* and assigns the values to the proper variables. This subroutine also determines the layer structure for a single, isolated module and stores the information in *trg* (layer thicknesses) and *v0* (conduction band edge energy in J).

deg

Applies the Pauli exclusion principle after a scattering event has taken place, and updates the electron distribution function *fde* if the transition is allowed.

`degin`

Initialization of the electron distribution function *fde* and the associated energy and momentum grid.

`df`

Calculation of the electron distribution function. The results are written to *edistr.dat.*

`dmass`

Calculates the position dependent effective electron mass across the structure.

`eei`

Calculation of the final state of the electrons involved in intersubband e-e scattering. The input/output parameters are:

e,kx,ky: energy and wave vector of first particle

n,m: initial and final subband of first particle

e1,kx1,ky1,n1: similar parameters for the partner electron

`ee`

Calculation of the final state of the electrons involved in intrasubband e-e scattering. The input/output parameters are:

e,kx,ky: energy and wave vector of first particle

n,m: initial and final subband of first particle

e1,kx1,ky1,n1: similar parameters for the partner electron

`eisubsc`

Calculation of the band structure and wavefunctions. The full potential profile of three modules and their associated wavefunctions *psi* are determined using the subroutine *schroed.* This subroutine also calculates the tight-binding elements and exports the results in *deltae.dat*. To speed up the calculation of the scattering matrix elements, *psi* is interpolated on a sparser grid, resulting in *zeta*.

`fract`

Determines the fraction of carriers in each subband.

`gaas`

Initialization of the material parameters of GaAs, such as the effective mass, dielectric constant and LO-phonon energy, according to Adachi.

`geneh`

Determination of the maximum scattering rates for intra- and intersubband e-e scattering, used in the Monte Carlo simulation.

`impur`

Calculation and rejection of the final state of e-imp scattering. Input parameters are:

e: energy of the scattered electron

kx,ky: momentum of the scattered electron

n: initial subband

m: final subband

iv: band to which the particle belongs, always equal to 1 (conduction band)

part_wt: particle weight, i.e. a weighting factor to determine how much of the density matrix describing the electron is scattered

`intrough`

Calculation and rejection of the final state of e-interface roughness scattering. Input parameters are:

e: energy of the scattered electron

kx,ky: momentum of the scattered electron

n: initial subband

m: final subband

iv: band to which the particle belongs, always equal to 1 (conduction band)

part_wt: particle weight, i.e. a weighting factor to determine how much of the density matrix describing the electron is scattered

`initfd`

Initialization of the electron distribution before the main algorithm commences. Electrons are randomly assigned to a subband and a momentum based on a thermal distribution.

`integr`

Integrates a function f using the Simpson rule.

`maxff`

Finds and returns the maximum e-LO-phonon form factor for any given set of initial and final subbands. This information is used in the subroutines *polabs* and *polems*, and in the determination of a maximum LO-phonon scattering rate used in the MC simulation.

`mkpot`

Subroutine called by *eisubsc*, used to build the potential profile of a QC structure under bias and return it in $v2$.

`phopho`

Implements the finite lifetime *tauph* of the LO-phonon population and updates the hot LO-phonon population *phd* accordingly.

`phsc`

This subroutine calculates the maximum scattering rates for e-imp and e-LO-phonon scattering, using the scattering form factors found from *aqnm*. The resulting cumulative rates are stored in $scr(i, n, m, k, ibd)$ according to the scattering electron's initial energy index i, initial and final subband n and m, band $ibd = 1$ and scattering mechanism k:

k=1: LO-phonon absorption

k=2: LO-phonon emission

k=3: acoustic phonon scattering

k=4: impurity scattering

k=5: interface roughness scattering

`polabs`

Calculation and rejection of the final state of e-LO-phonon absorption. Input parameters are:

e: energy of the scattered electron

kx,ky: momentum of the scattered electron

n: initial subband

m: final subband

iv: band to which the particle belongs, always equal to 1 (conduction band)

part_wt: particle weight, i.e. a weighting factor to determine how much of the density matrix describing the electron is scattered

`polems`

Calculation and rejection of the final state of e-LO-phonon emission. Input parameters are:

e: energy of the scattered electron

kx,ky: momentum of the scattered electron

n: initial subband

m: final subband

iv: band to which the particle belongs, always equal to 1 (conduction band)

part_wt: particle weight, i.e. a weighting factor to determine how much of the density matrix describing the electron is scattered

`thimp`

Integration of the e-imp form factor over 2π, used in the determination of the maximum scattering rates in *phsc*.

thir

Integration of the e-interface roughness form factor over 2π, used in the determination of the maximum scattering rates in *phsc*.

thint

Integration of the e-LO-phonon form factor over 2π, used in the determination of the maximum scattering rates in *phsc*.

volo

This subroutine chooses a scattering mechanism and final subband by generating a random number and comparing it to the maximum scattering rate. The input/output parameters are:

e: electron energy
n: initial subband
itype: scattering mechanism (explained in the discussion of *phsc*)
mtype: final subband
ntsm: maximum number of scattering mechanisms
iv,ibd2: initial and final band (obsolete, both are equal to 1 by default)

zpts

Generates wavefunctions *zeta* on a sparser grid than is the case for *psi*. This subroutine is called by *eisubsc*.

inirand

Initialization of the random number generator.

schroed

Schrödinger solver, using the shooting method implemented in *shoot*, *shootrec* and *shootpsi*. The obtained wavefunctions and energies are recorded in the output file *wf.dat*.

selecpsi

Sorts the wavefunctions by energy and by module, and stores the results in *submod,ns* and *ninmod* and the file *subbands.dat*. There is the possibility to crop wavefunctions within their module, as a first approximation to limit parasitic interaction between modules. This subroutine is called from *eisubsc*.

calcpi

This subroutine calculates the non-interacting polarizability Π based on the description by Maldague [28] and Lee and Galbraith [64]. The screened matrix elements for e-imp and e-e scattering are determined and stored in *vimscr* and *vscr*, respectively.

calcliouville

Determines the Liouville matrix that governs the coherent time evolution and stores it in *liouv*.

initdm

Initialization of all density matrices *dm* at the beginning of the simulation.

propagatedm

Propagates the density matrix of a particle over a free-flight time and updates the contents of *dm*.

collapse_dm

Adjusts the density matrix after a scattering event took away *part_wt* from subband *isub* in particle *ni*.

calcdm

Calculates the coherent evolution of the density matrix over 50 fs for the initial value $\rho_{ij} = 1$ and $\rho_{nm} = 0, (n \neq i \text{and} m \neq j)$. The results are stored in *dmtempl* and used to calculate the coherent time evolution of electrons during free flight.

kbins_dm

Assigns the particles to bins based on their position in momentum space, and calculates a new density matrix with averaged momentum and energy to represent each bin. This subroutine reduces the total number of particles in the simulation, but preserves the total population.

generate_dm

Generates a new particle described by the density matrix with the specified parameters.

shuffle

Places the contents of the last particle with index *nelts* in particle *ni_i* and decrease the number of particles *nelts* by one.

A.2 Non-equilibrium Green's Function Simulation

To set up the NEGF simulations, I use the Wannier functions generated by Prof. Wacker's code. The input file used is *band1.inp*, which contains the information about the layer structure. The output files from *band.out* are:

wavef.dat: Wannier functions

ueberlapp.dat: wavefunction overlap integrals

pot.dat: conduction band profile for 7 modules (no bias)

kopplung.dat: interaction matrix elements T_1

The NEGF code consists of four files, *green10.F, subaqee.F, subdefreg.F* and *subeisubsc.F* and can be compiled with the instruction: `f77 -cpu:p7 -W -O1 -lg2c -llapack green10.F subaqee.F subdefreg.F subeisubsc.F` The output file is *a.out*. In the following sections, I will briefly summarize the different subroutines in the NEGF code.

aqnm

Determination of the form factors for e-imp and e-LO-phonon scattering. Output files are *ffim.dat* (e-imp) and *ffpop.dat* (e-LO-phonon scattering).

calc_GR

Implementation of the Dyson and Keldysh equations to calculate the Green's functions G^{ret} and $G^{<}$. The diagonal and off-diagonal blocks of the functions are stored separately.

calc_SR

Calculation of Σ^{ret} and $\Sigma^{<}$ from G^{ret} and $G^{<}$. This subroutine also performs the self-energy calculations needed to determine the scattering current after the algorithm has converged.

cgain

This subroutine sets up the gain calculations and determines the emission energy for which the gain is calculated.

calcpi

This subroutine calculates the non-interacting polarizability Π based on the description by Maldague [28] and Lee and Galbraith [64]. The screened matrix elements for e-imp and e-e scattering are determined and stored in *vimscr* and *vscr*, respectively.

calcfftheta

Calculation of the integrated form factors B used in the subroutine *calc_SR*.

findpop

Determines the subband population *pop*.

findtemp

Determines the subband temperature *temp* by assuming thermalized subband populations.

gaas

Initialization of the material parameters of GaAs, such as the effective mass, dielectric constant and LO-phonon energy, according to Adachi.

`initgreen`

Initialization of the verious Green's functions and self-energies. Thermalized subbands are assumed with the population distributed equally across all subbands.

`inithamilt`

Initialization of the subband energies Er and interaction energy matrix U.

`fullmatrix`

This subroutine constructs the Green's function or self-energy for a system of 3 modules from its representative diagonal and off-diagonal elements.

`calc_dmat`

Calculation of the density matrix and electron distribution functions from $G^{<}$.

`normalize`

Normalization of the total population after each iteration of the algorithm.

`calc_dSR`

Calculation of $d\Sigma^{\mathrm{ret}}$,$d\Sigma^{\mathrm{adv}}$ and $d\Sigma^{<}$ from dG^{ret},dG^{adv} and $dG^{<}$. These functions are used for the calculation of the gain.

`calc_dGR`

Calculation of dG^{ret},dG^{adv} and $dG^{<}$ using the formalism described in 4.6.

Bibliography

[1] S. C. Lee and A. Wacker. Theoretical analysis of spectral gain in a terahertz quantum-cascade laser: Prospects for gain at 1 thz. *Appl. Phys. Lett.*, 83:2506, 2003.

[2] J.S. Blakemore. Semiconducting and other major properties of gallium arsenide. *J. Appl. Phys.*, 53:R123–R179, 1982.

[3] M. Tacke. New developments and applications of tunable IR lead salt lasers. *Infrared Phys.Technol.*, 36:447–463, 1995.

[4] S.E. Rosenbaum, B.K. Kormanyos, L.M. Jelloin, M. Matloubian, A.S. Brown, E. Larson, L.D. Nguyen, M.A. Thompson, L.P.B. Katehi, and G.M. Rebeiz. 155- and 213-Ghz AlInAs/GaInAs/InP HEMT MMIC oscillators. *IEEE Trans.Microwave Theory Tech.*, 43:927–932, 1995.

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

[6] R.F. Kazarinov and R.A. Suris. Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice. *Soviet Physics: Semiconductors*, 5:707–709, 1971.

[7] 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.

[8] R. Colombelli, F. Capasso, C. Gmachl, A.L. Hutchinson, D.L. Sivco, A. Tredicucci, M.C. Wanke, A.M. Sergent, and A.Y. Cho. Far-infrared surface-

plasmon quantum-cascade lasers at 21.5 μm and 24 μm wavelengths. *Appl. Phys. Lett.*, 78:2620–2622, 2001.

[9] R. Köhler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davis, D. A. Ritchie, R. C. Iotti, and F. Rossi. Terahertz semiconductor-heterostructure laser. *Nature*, 417:156, 2002.

[10] J. Faist, F. Capasso, D.L. Sivco, A.L. Hutchinson, S.N.G. Chu, and A.Y. Cho. Short wavelength ($\lambda \sim 3.4$ μm) quantum cascade laser based on strained compensated ingaas/alinas. *Appl. Phys. Lett.*, 92:680–682, 1998.

[11] 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.

[12] H. Callebaut, S. Kumar, B.S. Williams, Q. Hu, and J.L. Reno. Analysis of transport properties of terahertz quantum cascade lasers. *Appl. Phys. Lett.*, 83:207, 2003.

[13] R. Köhler, R. C. Iotti, A. Tredicucci, and F. Rossi. *Appl. Phys. Lett.*, 79, 2001.

[14] F. Compagnone, A. Di Carlo, and P. Lugli. Monte carlo simulation of electron dynamics in superlattice quantum cascade lasers. *Appl. Phys. Lett.*, 80, 2002.

[15] D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić. Mechanisms of temperature performance degradation in terahertz quantum-cascade lasers. *Appl. Phys. Lett.*, 82, 2003.

[16] J. Faist, F. Capasso, C. Sirtori, D. L. Sivco, A. L. Hutchinson, S. N. G. Chu, and A. Y. Cho. Narrowing of the intersubband electroluminescent spectrum in coupled-quantum-well heterostructures. *Appl. Phys. Lett.*, 65, 1994.

[17] H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka, and T. Matsusue. Interface roughness scattering in gaas/alas quantum wells. *Appl. Phys. Lett.*, 51, 1987.

[18] S. M. Goodnick and P. Lugli. *Appl. Phys. Lett.*, 51, 1987.

[19] M. Dür, S. M. Goodnick, and P. Lugli. Monte carlo simulation of intersubband relaxation in wide, uniformly doped gaas/al$_x$ga$_{1-x}$as quantum wells. *Phys. Rev. B*, 54, 1996.

[20] H. Vaupel, P. Thomas, O. Kuehn, V. May, K. Maschke, A. P. Heberle, W. W. Ruehle, and R. Koehler. Dissipative tunneling in asymmetric double-quantum-well systems: A coherence phenomenon. *Phys. Rev. B*, 53:16531, 1996.

[21] F. Eickemeyer, K. Reimann, M. Woerner, T. Elsaesser, S. Barbieri, C. Sirtori, G. Strasser, T. Mueller, R. Bratschitsch, and K. Unterrainer. Ultrafast coherent electron transport in semiconductor quantum cascade structures. *Phys. Rev. Lett.*, 89:047402, 2002.

[22] C. W. Luo, K. Reimann, M. Woerner, T. Elsaesser, R. Hey, and K. H. Ploog. Rabi oscillations of intersubband transitions in gaas/algaas mqws. *Semicond. Sci. Technol.*, 19:S285, 2004.

[23] H. Callebaut, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno. Importance of electron-impurity scattering for electron transport in terahertz quantum-cascade lasers. *Appl. Phys. Lett.*, 84:645, 2004.

[24] H. Willenberg, G.H. Doehler, and J. Faist. Intersubband gain in a bloch oscillator and quantum cascade laser. *Phys. Rev. B*, 67:085315, 2003.

[25] G. Mahan. *Many-particle Physics.* Plenum, New York, 1990.

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

[27] C.J. Hearn. *The physics of nonlinear transport in semiconductors, edited by D.K. Ferry, J.R. Barker and C. Jacoboni*, pages 153–166. Plenum, New York, 1980.

[28] P.F. Maldague. Many-body corrections to the polarizability of the two-dimensional electron gas. *Surf.Sci.*, 73:296, 1978.

[29] T. Unuma, M. Yoshita, T. Noda, H. Sakaki, and H. Akiyama. Intersubband absorption linewidth in gaas quantum wells due to scattering by interface roughness, phonons, alloy disorder, and impurities. *J. Appl. Phys.*, 93, 2003.

[30] S. Luryi. Frequency limit of double-barrier resonant-tunneling oscillators. *Appl. Phys. Lett.*, 47:490, 1985.

[31] A. Mircetic, D. Indjin, Z. Ikonic, P. Harrison, V. Milanovic, and R.W. Kelsall. Towards automated design of quantum cascade lasers. *J. Appl. Phys.*, 97:84506, 2005.

[32] R. Koehler, R. C. Iotti, A. Tredicucci, and F. Rossi. Design and simulation of terahertz quantum cascade lasers. *Appl. Phys. Lett.*, 79:3920, 2001.

[33] B.S. Williams, S. Kumar, H. Callebaut, Q. Hu, and J.L. Reno. Terahertz quantum-cascade laser operating up to 137 k. *Appl. Phys. Lett.*, 83:5142, 2003.

[34] S. Mukamel. *Principles of nonlinear optical spectroscopy.* Oxford University Press, New York, Oxford, 1995.

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

[36] 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.

[37] 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.*, 34:1722–1729, 1998.

[38] B. S. Williams. Terahertz quantum cascade lasers. *M.I.T. Doctoral thesis*, 2003.

[39] T. Kuhn and F. Rossi. Monte carlo simulation of ultrafast processes in photoexcited semiconductors: Coherent and incoherent dynamics. *Phys. Rev. B*, 46:7496, 1992.

[40] L.P. Kadanoff and G. Baym. *Quantum Statistical Mechanics.* Benjamin, New York, 1962.

[41] L. V. Keldysh. *Sov. Phys. JETP*, 20:1018, 1965.

[42] R. Mattuck. *A guide to Feynman diagrams in the many-body problem.* McGraw-Hill Inc., New York, 1976.

[43] T. Kuhn. *Theory of Transport Properties of Semiconductor Nanostructures (ed. by E. Schoell).* Chapman and Hall, London, 1998.

[44] G. Mahan. *Many Particle Physics.* Kluwer Academic, New York, 2000.

[45] S. Datta. *Electronic transport in Mesoscopic Systems.* Cambridge University Press, New York, 1995.

[46] A. Wacker. Semiconductor superlattices: a model system for nonlinear transport. *Phys. Rep.*, 357:1–111, 2002.

[47] A. Wacker. Transport in nanostructures: A comparison between nonequilibrium green functions and density matrices. *Adv. in Solid State Phys.*, 41:199–210, 2001.

[48] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic. Single and multiband modeling of quantum electron transport through layered semiconductor devices. *J. Appl. Phys.*, 81:7845, 1997.

[49] W.W. Chow and S. W. Koch. *Semiconductor Laser Fundamentals.* Springer, Berlin, 1999.

[50] H. Haug. *Quantum Theory of the Optical and Electronic Properties of Semiconductors.* World Scientific, Singapore, 2004.

[51] S. C. Lee and A. Wacker. Nonequilibrium greens function theory for transport and gain properties of quantum cascade structures. *Phys. Rev. B*, 66, 2002.

[52] S.-C. Lee, F. Banit, M. Woerner, and A. Wacker. Quantum-mechanical wavepacket transport in quantum cascade laser structures. *Cond. Mat.*, 7:0507410, 2005.

[53] F. Banit, S.-C. Lee, A. Knorr, and A. Wacker. Self-consistent theory of the gain linewidth for quantum-cascade lasers. *Appl. Phys. Lett.*, 86:041108, 2005.

[54] W. Kohn. Analytic properties of bloch waves and wannier functions. *Phys. Rev.*, 115:809, 1959.

[55] G. H. Wannier. The structure of electronic excitation levels in insulating crystals. *Phys. Rev.*, 52:191, 1937.

[56] B. S. Williams, H. Callebaut, Q. Hu, and J. L. Reno. Magnetotunneling spectroscopy of resonant anticrossing in terahertz intersubband emitters. *Appl. Phys. Lett.*, 79:4444, 2001.

[57] S. Kumar, B. S. Williams, S. Kohen, Q. Hu, and J. L. Reno. Continuous-wave operation of terahertz quantum-cascade lasers above liquid-nitrogen temperature. *Appl. Phys. Lett.*, 84:2494, 2004.

[58] F. Banit, S. C. Lee, A. Knorr, and A. Wacker. Self-consistent theory of the gain linewidth for quantum-cascade lasers. *Appl. Phys. Lett.*, 86:041108, 2005.

[59] S. Kohen, B.S. Williams, and Q. Hu. Electromagnetic modeling of terahertz quantum cascade laser waveguides and resonators. *J. Appl. Phys.*, 97:053106, 2005.

[60] B.S. Williams, S. Kumar, Q. Hu, and J.L. Reno. Resonant-phonon terahertz quantum-cascade laser operating at 2.1 thz (λ =141 μm). *Electron. Lett.*, 40:431, 2004.

[61] H. Callebaut, S. Kumar, B.S. Williams, Q. Hu, and J. L. Reno. Importance of electron-impurity scattering for electron transport in terahertz quantum-cascade lasers. *Appl. Phys. Lett.*, 84:645, 2004.

[62] T. Ando, A. B. Fowler, and F. Stern. Electronic properties of two-dimensional systems. *Rev. Mod. Phys.*, 54:437, 1982.

[63] B. S. Williams, S. Kumar, H. Callebaut, Q. Hu, and J. L. Reno. Terahertz quantum-cascade laser at $\lambda \approx 100$ μm using metal waveguide for mode confinement. *Appl. Phys. Lett.*, 83:2124, 2003.

[64] S.-C. Lee and I. Galbraith. Intersubband and intrasubband electronic scattering rates in semiconductor quantum wells. *Phys. Rev. B*, 59:15796, 1999.