Transport in quantum cascade lasers

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
ROMAIN LÉON TERAZZI
Diplôme de Physicien, Université de Neuchâtel
born on April 21, 1978
citizen of Genève, GE - Switzerland

accepted on the recommendation of
Prof. Dr. Jérôme Faist, examiner
Prof. Dr. Andreas Wacker, co-examiner
Prof. Dr. Andreas Vaterlaus, co-examiner

2012
Pour Himali et Lisa
Abstract

The simulation of transport in semiconductor heterostructures like quantum cascade lasers is of central interest as it enables the knowledge of the electrons dynamics inside such structures, allowing the determination of electrical and optical properties of the latter. These human-designed structures have an atomic resolution and therefore require a quantum mechanical description. The latter can be performed at different levels. The basic description gives the band-structure that represents where electrons can exist in the structure. However this description does not provide information about the transport properties, as the latter require the knowledge of the interaction of electrons with various sources of scattering inside the structure.

In a quantum mechanical description of transport, the coherent processes, where electrons wavepackets propagate freely along the energy-dispersion of structure, have a strong interplay with the incoherent processes where scattering destroys the phase relation between electrons. Even with very simple structures the dynamics of electrons in a quantized system with scattering is a complicated task. This is especially true for quantum cascade lasers and we have to perform approximations.

In this work we present an effective transport model that relies both on coherent and incoherent transport processes. We have retained from coherent effects the resonant tunneling, where electrons oscillate between two quantum states coupled by a potential barrier and give rise to a current inside the structure when scattering localize them on a side of the barrier. The role of incoherent scattering is then to redistribute the electrons between quantum states inside the structure.

The model represents the heterostructure by zero dimensional effective quantum states where scattering have been averaged while the current is modeled with resonant tunneling and accounts for second-order terms given by intra-subband transitions. An important aspect of the model is its lightness in term of computational effort. This has enabled the computation of the optical power emitted by the heterostructure in a good agreement with experiments. A numerical model for automatic optimization of structures is also
In lasers, the light amplification is caused by optical gain between pairs of quantum states. In this work we have also studied the gain mechanisms in quantum cascade lasers. We show measurements and simulations that give evidences of the existence of a second-order gain (bloch gain). The gain mechanism on which relies a Bloch oscillator have the same nature.
Résumé

La simulation du transport dans les hétérostructures semi-conductrices comme les lasers à cascades quantiques est fondamentale pour la compréhension de ces dispositifs opto-électroniques car elle permet de connaître la dynamique des électrons dans la structure et ainsi permet de calculer les propriétés optiques et électriques de ces dernières. Ces structures sont dessinées à une résolution atomique et nécessitent donc un traitement quantique. Ce dernier peut être effectué à différents niveaux. La description la plus simple donne la structure de bande qui renseigne sur les états d’énergie que peuvent occuper les électrons. Toutefois elle ne fournit pas d’information sur le transport car ce dernier nécessite de connaître les processus de diffusions dans la structure.

Lorsque ces systèmes sont décrits à l’aide de la mécanique quantique, les processus cohérents où les électrons se propagent librement dans la structure de bande sont interrompus par des processus incohérents qui détruisent la relation de phase entre les électrons. Même pour de très simple structures la dynamique des électrons dans un système quantifié est particulièrement complexe. Ceci est en particulier vrai pour les laser à cascades quantiques et certaines approximations doivent être envisagées.

Dans ce travail nous présentons un modèle de transport qui s’appuie sur des processus cohérents et incohérents. De la cohérence nous avons retenu l’effet tunnel résonant où les électrons oscillent entre deux états quantiques couplés par une barrière. Ce mécanisme donne naissance à un courant lorsque des électrons sont diffusés et dès lors localisés d’un côté de la barrière. Les processus incohérents redistribuent les électrons entre les états quantiques de la structure.

Dans notre modèle nous représentons l’hétérostructure par des états zéro-dimensionnels où les mécanismes de diffusion ont été moyennés. Toutefois, le courant est modélisé à l’aide de processus de second-ordre. Un aspect important du modèle est sa légèreté en terme de puissance de calcul. Ceci a permis de calculer la puissance optique émise par des structures laser. Un très bon accord est obtenu entre les prédictions théoriques et les mesures expérimentales. Une technique permettant d’optimiser automatiquement les
structures à cascade est présentée avec certains résultats prometteurs.

Dans les lasers, l’amplification du champ optique est due au gain optique entre des paires d’états quantiques. Dans le présent travail, nous avons aussi étudié les mécanismes de gain dans les lasers à cascade. Nous présentons des mesures et des simulations qui fournissent des prévues qu’un gain de second-ordre (Gain de Bloch) existe. Le mécanisme de gain sur lequel repose l’oscillateur de Bloch est le même que celui que nous avons observé.
# Contents

1 Introduction  
1.1 Transport regimes in quantum cascade lasers  3  
1.2 Organisation of the text  5  
1.2.1 Detailed structure of the document  6  

2 Band structure of bulk III-V semiconductors  9  
2.1 Lattice, Reciprocal lattice and Brillouin zones  9  
2.2 Chemical bonds, energy bands and semiconductors  12  
2.3 III-V semiconductors  13  
2.4 Calculation of the electronic states, the $k \cdot p$ analysis  15  
2.5 The Kane model and the conduction-band non-parabolicity  20  
2.5.1 Numerical results for various III-V alloys  25  
2.6 Three-band model: conduction, light-hole and split-off bands  26  
2.7 Effective two-band model  29  
2.8 Ternary and quaternary III-V alloys: the virtual crystal approximation  31  

3 Heterostructures and the envelop function approximation  35  
3.1 Multi-band model  38  
3.2 Envelop function formalism in the effective two-band model  42  
3.2.1 Calculation of dipole moments between conduction-subbands  45  

4 The band-structure of quantum cascade lasers  49
4.1 Formulation of the problem for energy-eigenstates ........................................ 49
4.2 Numerical implementation of the solution ......................................................... 52
  4.2.1 General explanation of the method ............................................................... 52
  4.2.2 Numerical solution with an electro-static potential ...................................... 57
  4.2.3 Field-related conditions and the continuum .................................................. 59
4.3 Band-structure of a quantum cascade laser ....................................................... 60
  4.3.1 Band-structure of the period at zero-field .................................................... 60
  4.3.2 Alignment electric field .................................................................................. 62
  4.3.3 Conceptual transport model ............................................................................ 62
  4.3.4 Injector selectivity .......................................................................................... 64
4.4 Tight-binding parameters, coupling energies ...................................................... 65
  4.4.1 Tight-binding: statement of the problem ....................................................... 65
  4.4.2 Calculation of the coupling energies ............................................................... 65
  4.4.3 Evaluation of $K_{ij}$ and $T_{ij}$ ...................................................................... 69
  4.4.4 Numerical examples ....................................................................................... 73
4.5 Self-consistent potential .................................................................................... 75
  4.5.1 Modeling the self-consistent potential ............................................................ 76
  4.5.2 Periodicity of the charge density ..................................................................... 78
  4.5.3 Example of self-consistent potential .............................................................. 79
  4.5.4 Modeling issues ............................................................................................. 80
  4.5.5 Self-consistent potential from a thermal distribution ...................................... 81
  4.5.6 Self-consistent solution and convergence ...................................................... 85

5 Non-radiative scattering between subbands ............................................................ 89
  5.1 Definition of $\Gamma_{\text{inter}}$ and $\Gamma_{\text{intra}}$ ......................................................... 89
    5.1.1 Definition of $\Gamma_{\text{inter}}$ ........................................................................... 91
    5.1.2 Definition of $\Gamma_{\text{intra}}$ ........................................................................... 91
  5.2 Scattering rates, dephasing times and optical linewidths .................................... 92
5.2.1 Calculation of the inter-subband scattering rates ......................... 93
5.2.2 Calculation of the dephasing times ........................................... 94
5.2.3 Calculation of the optical linewidth .......................................... 95

5.3 Elastic scattering mechanisms ..................................................... 95
5.3.1 Interface roughness .............................................................. 95
5.3.2 Alloy disorder ........................................................................ 103
5.3.3 Ionized impurities scattering ................................................... 106

5.4 Inelastic non-radiative scattering mechanisms: LO-phonon scattering ... 109

5.5 Typical scattering rates ............................................................... 113

5.6 Other sources of scattering .......................................................... 115

6 Basic transport models and requirements ........................................... 117

6.1 Derivation and general aspects of rate equations ............................... 118
6.1.1 Rate equations from the Boltzmann equations in a superlattice ....... 118
6.1.2 Subbands with thermal distributions ........................................... 119
6.1.3 Factorization of the distribution ............................................... 121
6.1.4 Isolated subbands: matrix formulation and solution ..................... 122
6.1.5 Inhomogeneous rate equations .................................................. 123
6.1.6 Rate equations with hidden states ............................................. 125

6.2 Simple rate equations in a quantum cascade laser ............................ 126

6.3 Rate equations with boundary conditions ....................................... 131
6.3.1 Periodicity of the spectrum in a quantum cascade structure .......... 131
6.3.2 Computation of the wavefunctions .......................................... 134
6.3.3 Rate equations with boundary conditions .................................. 137

6.4 Modeling transport across a barrier .............................................. 144
6.4.1 Basic model: two coupled states across a barrier potential .......... 144
6.4.2 Current damping and the density-matrix formalism ..................... 147
6.4.3 The density matrix ............................................................... 149
6.4.4 Two-state system treated with the density matrix ....................... 150
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.5</td>
<td>Two-state system with periodic boundary conditions</td>
<td>152</td>
</tr>
<tr>
<td>6.4.6</td>
<td>Resonant current formula and quantum cascade structures</td>
<td>157</td>
</tr>
<tr>
<td>6.4.7</td>
<td>Structure of the dephasing time $\tau_\perp$</td>
<td>159</td>
</tr>
<tr>
<td>6.5</td>
<td>A toy model for far-infrared single-quantum-well structure</td>
<td>162</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Calculation of the current density</td>
<td>164</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Analysis of the current</td>
<td>167</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Population inversion</td>
<td>169</td>
</tr>
<tr>
<td>6.6</td>
<td>Two subbands in the density-matrix formalism</td>
<td>170</td>
</tr>
<tr>
<td>6.6.1</td>
<td>Failure of the first order current model</td>
<td>171</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Current between a pair of subbands</td>
<td>175</td>
</tr>
<tr>
<td>6.6.3</td>
<td>Interpretation of the second-order current density</td>
<td>184</td>
</tr>
<tr>
<td>6.6.4</td>
<td>Second-order current in a 0-dimensional effective model</td>
<td>187</td>
</tr>
<tr>
<td>7</td>
<td>Towards an effective transport model</td>
<td>195</td>
</tr>
<tr>
<td>7.1</td>
<td>Formal derivation of an effective transport model</td>
<td>196</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Foreword on the basic physics of the model</td>
<td>196</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Density matrix revisited: Liouville operator and scattering</td>
<td>197</td>
</tr>
<tr>
<td>7.1.3</td>
<td>General structure of the scattering operators</td>
<td>201</td>
</tr>
<tr>
<td>7.1.4</td>
<td>Second-order currents in the effective Liouville model</td>
<td>205</td>
</tr>
<tr>
<td>7.1.5</td>
<td>Stationary solution: effective rate equations</td>
<td>206</td>
</tr>
<tr>
<td>7.1.6</td>
<td>Periodic systems with one coupling barrier</td>
<td>208</td>
</tr>
<tr>
<td>7.1.7</td>
<td>Systems with several coupling barriers</td>
<td>212</td>
</tr>
<tr>
<td>7.2</td>
<td>Importance of the electronic temperature</td>
<td>216</td>
</tr>
<tr>
<td>7.2.1</td>
<td>The kinetic energy balance</td>
<td>218</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Computational structure of the model</td>
<td>222</td>
</tr>
<tr>
<td>7.3</td>
<td>Current-voltage curves: model against experiment</td>
<td>222</td>
</tr>
<tr>
<td>7.4</td>
<td>Competing models</td>
<td>227</td>
</tr>
<tr>
<td>8</td>
<td>Light interaction modeling in quantum cascade lasers</td>
<td>231</td>
</tr>
<tr>
<td>8.1</td>
<td>Interaction between subbands and the electromagnetic field</td>
<td>232</td>
</tr>
</tbody>
</table>
8.1.1 The dipole approximation .................................. 232
8.1.2 Absorption coefficient of a heterostructure .................. 234
8.1.3 Selection rules for optical transitions in heterostructures .... 236

8.2 Gain mechanisms in heterostructures .......................... 239
8.2.1 First-order gain model ..................................... 240
8.2.2 Inhomogeneous broadening: application to first-order gain .... 242
8.2.3 Non-parabolicity effects on first-order gain .................... 245
8.2.4 Second-order gain .......................................... 248
8.2.5 Inhomogeneous broadening and second-order gain ............ 260
8.2.6 Non-parabolicity effects on second-order gain ................ 262
8.2.7 Second-order gain and the Bloch oscillator ................... 263
8.2.8 Bloch gain in a mid-infrared quantum cascade laser .......... 267
8.2.9 Refined transport simulations for N258 sample ................ 273

8.3 Computation of light-current curves in the far-infrared ......... 279
8.3.1 The photon-flux density .................................... 282
8.3.2 Impact of resonant tunneling on emitted power ............... 285

8.4 Coupled optical field intensity and effective transport model .... 290
8.4.1 Method for computing the photon-flux density $S$ ................ 290
8.4.2 Computation of the spectral gain ................................ 292
8.4.3 Light-current-voltage curves of a two-phonon structure ....... 293

8.5 Towards the automatic optimization of quantum cascade lasers ...... 297

9 Conclusions and perspectives .................................... 305

A An example of super-self-consistent computations .................. 311

B Dephasing times: comparisons between models ....................... 317

C Steady-state solutions with the Laplace average ..................... 321
Chapter 1

Introduction

This work is devoted to the modeling of the transport in quantum cascade lasers. The latter are semiconductor heterostructures where the electronic properties have been designed in order to achieve a population inversion between quantum states. The first demonstration was obtained by J. Faist et al. [30] with a structure emitting at 4.2 µm in the mid-infrared region of the spectrum. The idea of light amplification in semiconductor heterostructures was examined in the seminal work of R. F. Kazarinov and R. A. Suris [55, 56]. However they have considered superlattices [100] structures and the optical gain required for laser action was found in an unstable electrical regime [19]. In the quantum cascade lasers, passive semiconductor layers are inserted between the gain regions in order to stabilize the structure.

The complexity of the electronic structure is however greatly increased and the modeling of the dynamics of electrons require dedicated simulation techniques including various physical models. The latter may account more or less on the fundamental quantum nature of the transport mechanisms.

The statement of the problem can be formulated quite simply: what is the value of the current driven by the structure when a constant electric field is applied on it. This gives the basic characteristic of the heterostructure, the current-voltage curve. The goal of a simulation is indeed to obtain the better agreement between the measured current and the computed one. Moreover the simulation gives access to the distribution of electrons inside the structure and the corresponding spectral optical gain can be obtained and may enable the computation of the power emitted by the structure where the latter is embedded into an optical cavity.

Here, we only give a sketch of the various simulation techniques that are currently used to model quantum cascade structures; a more detailed discussion is given in (7.4). The
models can be classified in relation to the physical models they use. The semi-classical models \cite{11,14,15,24,45–49} are based on the Boltzmann equations where the scattering operator has a structure similar to rate equations. The basic quantum ingredients are the band-structure and the scattering rates. However we have to notice that some models \cite{15,36,49} have included coherent effects with resonant tunneling between a few states in the structure. These models require the density matrix formalism in order to treat the coherent effects together with the incoherent scattering. In contrast, the transport can be modeled with the non-equilibrium Green’s functions \cite{61,62,64,73,74,109,119}. In these models the problem is addressed in a completely quantum mechanical framework. The potential that gives rise to the band-structure and the potentials that cause scattering are treated on an equal footing. We can roughly say that the semi-classical models use the minimal knowledge of the quantum structure to obtain the better agreement with the experiment, while quantum models emphasize on the fundamental nature of the transport. Such models have enabled for example the understanding of the origin of the current in quantum cascade structures \cite{63} and in general may used for the modeling of new quantum effects in the transport.

In this work, we have developed a hybrid model that implements both scattering as the semi-classical models and simple quantum effects with resonant tunneling. The problem is formulated in the density matrix formalism \cite{81,96}. Semi-classical models frequently use a three-dimensional modeling of the transport and solutions are obtained by using Monte-Carlo methods \cite{51}. Here we have considered an effective model where the scattering rates are averaged \cite{24} and the transport occurs at the end between zero dimensional states. We however consider that resonant tunneling is driving the current in the structure by using the quantum mechanical expression of the current. This was demonstrated \cite{63} and we aimed to keep explicitly this result. Resonant tunneling effects were already examined \cite{14,49}, however tunneling was implemented to the first order by conserving the wave-vector. Here we have extended the model by implementing second-order results \cite{108,115}. The latter are crucial in the correct modeling of the current magnitude \cite{96}.

In a second step, we have modeled the interaction of the electrons with the intensity of a laser field. The numerical modeling of the coupling between the light intensity and the dynamics of electrons has enabled us to predict the optical power emitted by quantum cascade lasers \cite{94} with a good agreement with experiment. This model was then used successfully in the automatic optimization of mid-infrared structures. It is important to outline that even if the model implements the dynamics of electrons with effective zero-dimensional states, we have developed a kind of \textit{ab initio} model, as we do not have fitting parameters.
1.1. Transport regimes in quantum cascade lasers

The transport in a far-infrared laser was also investigated [83] in a simple four-state density matrix model. Special features in the light-current characteristic outlined the importance of resonant tunneling. They were predicted by the model and then observed experimentally.

Complementary way we have investigated the gain mechanisms in quantum cascade lasers. The interaction with the electromagnetic field that gives rise to the optical gain is dissociated from non-radiative scattering mechanisms in the first order model. In this model the optical gain is driven by the net population inversion as in atomic systems. H. Willemenberg et al. [115] and A. Wacker [109] have shown independently that the simultaneous treatment of optical transitions and scattering enables gain without a net population inversion. The gain in the Bloch oscillator is found to rely on the same mechanisms [10,115]. The predicted spectral gain has been observed experimentally in the mid-infrared [39,95] demonstrating the validity of the theoretical model. The theoretical curves were computed with a simple model and a good agreement is found for the shape of the gain. The current-voltage curves are well reproduced by our effective transport model. However the computation of the gain is more complicated as out-of-equilibrium electron distributions are required in order to obtain a good agreement with the experiment.

1.1 Transport regimes in quantum cascade lasers

The transport regime in a quantum cascade laser can be compared with the transport regimes in superlattices. The latter were examined in a very complete way [110,113]. The diagram in Fig.(1.1) is especially meaningful. The vertical axis represent the coupling energy $T_1$ between adjacent periods, while the horizontal axis represents the energy drop $qFd$ causes by an applied electric field $F$, where $d$ is the period length and $e = q_0$ is the signless elementary charge.

The shaded areas represent the validity range of various approximations. The energy $\Gamma$ is the typical broadening due to scattering. It is worth to remark that even for superlattices, the current in a transport regime where the typical energies $T_1$ and $eFd$ are of the same order of magnitude than $\Gamma$ cannot be calculated accurately by usual models.

The situation is even worse in quantum cascade lasers. If we consider the active period of a typical (two-phonon design) quantum cascade structure [9] as shown in Fig.(1.2). The active periods are coupled by a special barrier called the injection barrier. The carriers from the previous period (in state G) are injected into the upper laser-state 4. The transport regime for the injection process is sequential resonant tunneling and the adapted basis is the Wannier-basis [110] with states localized in a period.
Figure 1.1: The transport regimes in a superlattice are sketched. The condition $a \gg b$ on energies has been translated into $a > b$. This diagram was taken from [113].

Figure 1.2: Typical quantum cascade structure in the mid-infrared.
The electron injected in the upper laser-state (4) can decay by radiative or non-radiative processes to lower states 3, 2, 1 of the active region. Under an applied electric field the states in the active region are very close to Wannier-Stark states. Rigorously these states are rather resonances than bound-states as they are coupled to the continuum. However such coupling is avoided by the mini-gap formed by the injector region downstream and we usually approximate these states with bound-states. The leakage to the continuum being modeled by an escape time in rate equations.

The electrons in the lower states 3, 2, 1 are extracted by resonant tunneling to the injector region across an extraction barrier. The injector region is formed by a kind of superlattice structure. The transparency of this region is designed to allow transport by a mini-band and to avoid leakage by a mini-gap. The injector region is indeed finite and the term of mini-band/gap is inappropriate. However compared to the active region, the density of states is largely higher and this is the reason for using terms usually reserved to infinite systems. The role of the injector region is to thermalize the carriers in the injector ground-state and to prepare them to cascade to the next region. The transport in this region is typically incoherent and rely on pure scattering.

In this precise point we can realize the difference between semi-classical and quantum models. For the former, we impose incoherent transport as we do not have clear experimental clues that an electron in a lower state of the active region keep a phase relation with an electron injected from the injector ground-state to the next upper laser-state. The emission due to electrons of adjacent periods cannot be used from interferometry. In semi-classical models we therefore suppress completely the coherence by modeling transport with rate equations. In contrast in quantum models there is conceptually no need for imposing a particular transport regime as the model can treat any cases between the loss of coherence to coherent transport over a large spatial distance.

Our purpose here is not to describe quantitatively the transport in each region of a quantum cascade structure but to outline that quantum cascade structures reply on many different transport regimes. In this work we have tried to apply perturbative models adapted for each region in order to build an effective transport model.

1.2 Organisation of the text

We briefly sketch the organization of the body of the document. The latter is divided in ten sections. We first introduce the band structure of III-V semiconductors alloys. The envelop function is discussed and applied to quantum cascade structures. The non-radiative scattering mechanisms are developed. The transport models are first
introduced in (6) with scattering models. The resonant tunneling is introduced and an expression for the second-order current is obtained. The effective transport model is introduced formally (7) and compared to experiments. The coupling to the laser field is examined (8.3). The gain mechanisms, especially the second-order (bloch) gain, are discussed: measurements are shown and theoretical modeling developed. The emitted power of a two-phonon design is then compared to experimental results. Automatic optimization of quantum cascade structures is discussed (8.5) and some promising results are given. Conclusions and perspectives are given (9).

1.2.1 Detailed structure of the document

The band-structure in III-V semiconductors alloys is discussed in (2). We start from rather general considerations on the band-structure by recalling the $k \cdot p$ approximation (2.4) and the Kane model (2.5). The latter is reduced to an effective two-band model (2.7).

The envelop function approximation is then discussed (3) and applied specifically to the two-band model (3.2). We then discuss the calculation of the oscillator strength in heterostructures (3.2.1).

The band-structure of quantum cascade lasers is then addressed (4). The numerical methods are then discussed in some details (4.2). The alignment field of a quantum cascade structure is defined (4.3.2).

A tight-binding model is then formulated in order to compute the coupling energies across a barrier (4.4), the importance of non-parabolicity effects are outlined.

Finally the self-consistent potential in the Hartree approximation is introduced (4.5) with examples and some details about the algorithm considered in this work (4.5.6).

Non-radiative scattering sources are then discussed (5). The scattering rate, dephasing time and optical linewidth are defined (5.2). We then treat some scattering sources in details. The elastic scattering is first considered (5.3) with interface roughness (5.3.1) where a vertical correlation model is introduced (5.3.1). The relevance of the latter is then discussed in a practical example. The alloy disorder scattering is considered (5.3.2) as a limit case of the interface roughness scattering. Finally the scattering on ionized impurities is addressed (5.3.3), some details about the computation of the form-factors are given (5.3.3).

The main inelastic scattering source, the longitudinal optical phonon is addressed (5.4). The computation of the form-factor is given in details (5.4) as we have developed a kind of optimized algorithm in order to reduce the computational effort on this interaction.
Typical scattering rates are then computed for all examined interactions (5.5) a section is devoted to missing but relevant interactions (5.6).

We then begin the discussion on transport models (6) with basic rate equations (6.1). The latter are examined in details (6.1.4, 6.1.5, 6.1.6) as they are routinely used in rough modeling of the population inversion in quantum cascade lasers (6.2).

The boundary conditions on rate equations are introduced in (6.3) by discussing the periodicity of the spectrum in quantum cascade lasers (6.3.1). The specific topic of the computation of wavefunctions with an infinite coherence length is discussed (6.3.2) and the usual scattering models are introduced (6.3.3). The instabilities in the current-voltage curves are discussed (6.3.3) on a typical example.

The modeling of coherent transport across a barrier is discussed in details (6.4). A two-state problem is first considered (6.4.1) and the current damping in the usual Rabi formula is investigated (6.4.2). The density matrix formalism is introduced (6.4.3). The two-state problem is treated in the density matrix formalism (6.4.4) with periodic boundary conditions (6.4.5). The widely used formula for resonant tunneling in quantum cascade structures is obtained (6.4.6) and applied to a two-phonon design (6.4.6). The important parameters in resonant tunneling are outlined and a special attention is paid to the dephasing time (6.4.7).

A density matrix model is then formulated for a far-infrared structure (6.5). The structure of the current and the population invasion are analyzed in (6.5.2) and (6.5.3).

We then show the failure of the first order current expression (6.6) on two examples. The calculation of the second-order formula is then made in details (6.6.2). The solution is given in (6.6.2). The results are then interpreted in (6.6.3) and an effective zero dimensional model is obtained (6.6.4). The two previously examples of failure of the first-order current are revisited with the second-order expression (6.6.4). The importance of second-order current in a transport model for quantum cascade structures is outlined.

An effective transport model is discussed in (7) and formally addressed in (7.1). The master equation is introduced in (7.1.2) and the structure of scattering operators is dicussed in (7.1.3). The model in then reduced to effective rate equations (7.1.5). It is then implemented for one coupling barrier (7.1.6) and for severals barriers (7.1.7).

The importance of the electronic temperature is then discussed (7.2) and the computational structure of the model briefly sketched (7.2.2). The predictions of the model are then confronted to measurements (7.3) for mid-infrared structures.

The important section about competing models is found in (7.4). The interaction with the electromagnetic field is discussed in (8). The dipole approxi-
formation is obtained (8.1.1) and the absorption cross-section in heterostructure is defined (8.1.2) with selection rules (8.1.3).

The gain mechanisms in heterostructures are then discussed (8.2). The first order gain linewidth is obtained in different cases: the lifetime broadening is addressed in (8.2.1), the inhomogeneous broadening is included in (8.2.2) and the gain/absorption due to non-parabolicity is investigated in (8.2.3).

The second-order gain is introduced and discussed in (8.2.4). The effect of inhomogeneous broadening is addressed (8.2.5) and the effect of non-parabolicity is investigated (8.2.6). Experimental evidences for second-order gain in the mid-infrared are shown in (8.2.7). New transport simulations are presented in (8.2.9).

The interaction with a laser field is introduced in (8.3) for the four-state model in the far-infrared. The impact of resonant tunneling on the light-current characteristic is demonstrated in (8.3.2).

The effective transport model is then extended to the computation of the emitted power (8.4). The computation of the spectral gain is discussed (8.4.2) and the predictions are confronted to experiments (8.4.3): the limitations of the model are observed on the electronic temperature.

We finally consider automatic optimization of quantum cascade structures based on a kind of genetic algorithm (8.5) and we give some promising results.

The results are discussed and conclusions are given with some perspectives (9).
Chapter 2

Band structure of bulk III-V semiconductors

In this section, we review briefly the band structure of bulk semiconductors formed by III and V compounds like: InAs, InP, InSb, GaP, GaAs, GaSb, AlAs, AlSb, ...

We first give a short introduction to the calculation of electronic states in solids. Then we focus on semiconductors and the $\mathbf{k} \cdot \mathbf{p}$ approximation used to describe the electronic states near the edge of the conduction band. Afterwards we introduce the Kane model that allows to model in an elegant and efficient way the influence of the valence states by the introduction of an energy-dependent effective mass. At this point we will be ready to discuss an effective two-band model derived from the Kane Hamiltonian that will enable us to derive a simplified expression of the energy-dependent effective mass.

At the very end of the section we discuss briefly the virtual crystal approximation for ternary and quaternary alloys. The latter are very relevant for this work since the Ga$_{0.47}$In$_{0.53}$As / Al$_{0.48}$In$_{0.52}$As - lattice matched on InP is the canonical material system for mid-infrared quantum cascade lasers.

2.1 Lattice, Reciprocal lattice and Brillouin zones

We review here briefly the electronic states in crystalline solids. The reader should refer to textbooks for a complete introduction. [4,72]

A crystalline solid is by definition a macroscopic arrangement of atoms with a microscopic elementary cell that is repeated periodically. We can consider three basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ underlying the Bravais lattice. The latter is formed by every point $\mathbf{r}$ that can be reached
by a translation vector \( \mathbf{T} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3 \), where \( n_i, i = 1, 2, 3 \) are integers. This lattice is a mathematical network. The crystalline structure is obtained by associating at each point of the network an atomic neighborhood.

To the direct Bravais lattice is associated a \emph{reciprocal} lattice. The latter is naturally defined by considering a periodic function \( f(\mathbf{r}) \) of the Bravais lattice. The function can be expanded in a Fourier serie:

\[
 f(\mathbf{r}) = \sum_{\mathbf{G}} C_\mathbf{G} \exp(i \mathbf{G} \cdot \mathbf{r}).
 \]  

The periodicity of \( f \) under an arbitrary translation \( \mathbf{T} \) implies that \( \mathbf{G} \cdot \mathbf{T} = 2\pi p \) where \( p \) is an integer. If we express \( \mathbf{G} = h_1 \mathbf{A}_1 + h_2 \mathbf{A}_2 + h_3 \mathbf{A}_3 \) with \( h_i \) being integers, the basis vectors \( \mathbf{A}_i \) are defined.

In the reciprocal space (or \( k \)-space), the smallest volume centered at the origin and confined between \emph{Bragg planes} is called the first \emph{Brillouin zone}. It is sometimes called Wigner-Seitz cell. A Bragg plane is associated with a vector of the reciprocal lattice \( \mathbf{G} \). It is a plane perpendicular to \( \mathbf{G} \) that crosses the midpoint of \( \mathbf{G} \). It is defined by all the point reached by vectors \( \mathbf{k} \) satisfying:

\[
 \mathbf{k} \cdot \left( \frac{1}{2} \mathbf{G} \right) = \left( \frac{1}{2} \mathbf{G} \right)^2.
 \]

The Bragg planes enable to define higher order Brillouin zone. The latter are constituted by fragments of polyhedrons surrounding the first Brillouin zone. Any higher order Brillouin zone can be translated back to fill the first Brillouin zone with the help of reciprocal-space vectors \( \mathbf{G} \). Brillouin zones of all order provide a tiling of the reciprocal space. Therefore, any function in the reciprocal space can be expressed in the first Brillouin zone; although it becomes a multi-valued function. This mapping is very useful to discuss features of the band structure like band-extrema, degeneracy, band-gaps, etc.

In Fig. (2.1) a simple example of two-dimensional triangular lattice is shown. The first Brillouin zone can be further reduced by symmetry operations of the direct lattice to the \emph{irreducible} Brillouin zone. At each extremity of the latter, high symmetry points are defined.

The irreducible Brillouin zone can be easily obtained in a one-dimensional problem. The translational symmetry yield reduction of the reciprocal space to the first Brillouin zone \([-\pi/a, \pi/a]\) (\( a \) being the lattice constant) while the invariance under reflection reduces the zone to \([0, \pi/a]\) which cannot be further reduced. The irreducible Brillouin zone therefore directly depends of the symmetry of the crystalline structure.

The high symmetry points are usually defined using some conventions. For example in the triangular lattice: \( M \) denotes the mid-point of an edge, while \( K \) is a corner point.

By convention, the \( \Gamma \) point is always the center (origin) of the first Brillouin zone.
Figure 2.1: A triangular lattice is examined. (a) The direct lattice is shown, with basis vectors $a_1 = \frac{a}{2} (-1, \sqrt{3})$ and $a_2 = \frac{a}{2} (1, \sqrt{3})$, $a$ being the lattice constant. The unit cell is in shaded blue. (b) The reciprocal lattice is shown with basis vectors $G_1 = \frac{2\pi}{a} (-1, -\frac{1}{\sqrt{3}})$ and $G_2 = \frac{2\pi}{a} (1, \frac{1}{\sqrt{3}})$. The unit cell is in shaded blue, while the first Brillouin zone is in pink. The latter is found hexagonal. (c) The first Brillouin zone is shown with dashed lines showing the assumed underlying symmetry operations: rotations, reflections, etc. The high symmetry points are shown. In the reciprocal basis they read: $\Gamma = (0,0)$, $M = (\frac{1}{3}, 0)$ and $K = (\frac{1}{3}, -\frac{1}{3})$. A representation of the irreducible Brillouin zone is shown in shaded orange. (d) The second Brillouin zone is translated back into the first. (e) The third Brillouin zone is translated back into the first.
2.2 Chemical bonds, energy bands and semiconductors

The electronic states in solids arise basically from chemical bonds between atoms in the crystalline structure. As this topic is very large we will only give basic concepts.

An appropriated example is the covalent bond. The latter arises by the sharing of outer electrons between atoms. As we will discuss later it is relevant for III-V semiconductors. The covalent bond is highly directional due to the symmetry of the atomic wavefunctions: $s$, $p$, ...

For example in the hydrogen molecule $H_2$, the two $s$-electrons interact by a coulomb potential. As in the case of a two-level system, the states of the molecule (the coupled system) are composed by linear combinations of isolated 1s states (in the viewpoint of the tight-binding model). This gives rise to a low-energy state, the bonding state and a high-energy state the anti-bonding state.

In a crystalline solid, these molecular bonds give rise to energy bands: bonding bands and anti-bonding bands. The type of bonds between atoms (covalent, ionic, metallic, ...) determine the band-structure of the crystal, while the number of outer electrons (contributing to the chemical bonds) determine the filling of the bands.

Three elementary kinds of crystalline solids can be defined by their band-structure and the position of their Fermi level. The latter being, at absolute zero temperature, the energy up to which the bands are filled with electrons according to Pauli’s exclusion principle.

These cases are illustrated in Fig. (2.2). The band following the last completely filled bands (called the valence bands) is called the conduction band. As only the electrons in a partially filled band contribute to the electrical conduction, the types of crystalline solids are related to their electrical conductivity.

The metals are conductors as they have a partially filled conduction band. The insulators have an empty conduction band separated from the last valence bands by a large energy-gap (several electron-volts).

We consider the intermediate case of the semiconductors. Even if intrinsic (without supplementary electrons given by doping) semiconductors are insulators at cryogenic temperatures, the thermal spreading of the Fermi-Dirac distribution enables them to conduct electrical current at higher temperatures.
2.3 III-V semiconductors

In this work we will consider structures made of III-V semiconductors. Some examples are: InAs, InP, InSb, GaP, GaAs, GaSb, AlAs, AlSb. In the next paragraph we summarize briefly some of their properties. [8]

Their crystallize in the *Zinc-Blende* structure: two-interpenetrating face-centered cubic lattices. The latter is not a Bravais lattice as it have two atoms in the elementary cell, one at the origin and the other at \( \left( \frac{a}{4}, \frac{a}{4}, \frac{a}{4} \right) \). For example, one of In and the other of As. The crystalline structure is shown in Fig.(2.3a). The reciprocal lattice is a body centered cubic lattice. The first Brillouin zone is a truncated octahedron. The latter is show in Fig.(2.3b) with high-symmetry points (\( \Gamma, X, L, \text{ etc.} \)) with the following convention:
Figure 2.3: (a) The crystalline structure of III-V semiconductors is shown with two types of atoms A and B. (b) The first Brillouin zone is shown with high symmetry points and crystallographic axes. Both figures were taken from [8].

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Γ</td>
<td>Center of the Brillouin zone</td>
</tr>
<tr>
<td>K</td>
<td>Middle of an edge joining two hexagonal faces</td>
</tr>
<tr>
<td>L</td>
<td>Center of a hexagonal face</td>
</tr>
<tr>
<td>X</td>
<td>Center of a square face</td>
</tr>
</tbody>
</table>

The III-V compounds have 8 outer electrons. The covalent hybridization between Ga and As (for example) forms tetrahedral bonds. The s-type bonds are deeply bound and filled by 2 electrons. The remaining 6 electrons fill the 3 bonding p-type orbitals that give rise to valence bands. The anti-bonding orbitals are empty. The lowest lying s-type anti-bonding bond give rise to the conduction band. In III-V semiconductor, the top of the valence band is located at the Γ point. In such semi-conductors the spin-orbit coupling cannot be neglected and contributes by splitting valence bands and altering the effective mass, as it will be discussed in a few paragraphs.
2.4 Calculation of the electronic states, the \( k \cdot p \) analysis

Figure 2.4: (a) The bandstructure, of InAs is shown. The extrema of the topmost valence bands and the lowest-lying conduction band is located at \( \Gamma \) point, thus forming a direct gap. The gap energy is shown in shaded orange. (b) The bandstructure of Si is shown. The band-gap is indirect as the top of the valences band is at \( \Gamma \), while the lowest point of the conduction band is near X point. The basic figures are taken from [80].

Without giving more details we show the band-structure of bulk InAs in Fig.(2.4). Various kind of methods are used for the computation of the band-structure. The band-structure shown here was computed with the pseudo-potential method.

2.4 Calculation of the electronic states, the \( k \cdot p \) analysis

The equation that determines the electronic states in a crystalline solid is the stationary Schrödinger equation with a periodic potential \( V(\mathbf{r}) \):

\[
H \psi(\mathbf{r}) = \varepsilon \psi(\mathbf{r}) \quad \text{with} \quad H = \frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}).
\]  

Here \( m_0 \) is the mass of the bare electron. The potential \( V(\mathbf{r}) \) that models the electro-static potential felt by an electron in the crystalline structure has the periodicity of the Bravais lattice: \( V(\mathbf{r} + \mathbf{T}) = V(\mathbf{r}) \), where \( \mathbf{T} \) is an arbitrary translation of the direct lattice. It is worth to mention that the eigen-energies (the bands) are solutions of a one-electron
potential $V$. The latter arises from the atomic structure of the crystal, which is a many-electrons problem. We are implicitly assuming that the potential is averaged. The precise physical model used for the computation of this potential is out of the scope of this work. We therefore assume the existence of a periodic crystalline potential without giving more explanations.

As $H$ is periodic in $r$, the solutions have to be eigen-functions of the translation operator $T_T$. The translation operator acts on wavefunctions as: $T_T \psi(r) = \psi(r + T)$. Eigenfunctions satisfy:

$$T_T \psi(r) = e^{i k \cdot T} \psi(r).$$  \hfill (2.3)

The wavefunctions are therefore Bloch waves:

$$\psi_{n,k}(r) = N u_{n,k}(r) e^{i k \cdot r},$$  \hfill (2.4)

where $N$ is a normalization constant. The functions $u_{n,k}(r)$ are periodic functions of the Bravais lattice. The index $n$ is the band-index, while $k$ is a vector of the reciprocal space. The quantity $\hbar k$ is the crystal momentum. The latter is not the momentum of the electron as $\psi_{n,k}(r)$ cannot be an eigen-function of $p$ and $H$ simultaneously. If we let $p$ (the atomic momentum) acts on Bloch waves, it yields:

$$p \psi_{n,k} = e^{i k \cdot r} \hbar k u_{n,k} + e^{i k \cdot r} p u_{n,k} = e^{i k \cdot r} (\hbar k + p) u_{n,k}$$  \hfill (2.5)

We therefore have:

$$p^2 \psi_{n,k} = e^{i k \cdot r} (\hbar k + p)^2 u_{n,k}$$  \hfill (2.6)

The allowed values of $k$ depend on the macroscopic boundary conditions on the crystal. If $N_i$ is the number of elementary cells in the $i$-direction, the wavevector $k$ is quantified by the periodicity condition: $k \cdot (N_i a_i) = 2 \pi p_i$ where $p_i$ is an integer. For large $N_i$, the allowed values of $k_i$ form a quasi-continuum. The latter is usually approximated by the continuum limit.

We have anticipated the solution of the Schrödinger equation by adding to the solution the integer band-index $n$. The central equation can be obtained by inserting the wavefunction $\psi(r)$ expanded in plane waves together with the Fourier expansion of the potential $V(r)$ into Eq.(2.2). The central equation for a given $k$ yields a secular equation that gives the band quantization.

We are not interested here in the computation of the global band-structure shown in Fig.(2.4). The latter does not provide explicitly the relevant quantities in semiconductor physics: the effective mass and the wavefunctions. We only need to know the dispersion
2.4. Calculation of the electronic states, the $k \cdot p$ analysis

relation $\varepsilon_{nk}$ (the allowed energies) for a small range of $k$ values around the band extrema. This is given by the $k \cdot p$ analysis.

The periodic functions satisfy the differential equations: $u_{nk}(r)$:

$$\left( \frac{p^2}{2m_0} + V(r) + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar k \cdot p}{m_0} \right) u_{nk} = \varepsilon_{nk} u_{nk}(r) \tag{2.7}$$

This system can be formally rewritten in a form which is best suited for a perturbative expansion around the band extrema:

$$(H(k=0) + W(k)) u_{nk}(r) = \varepsilon_{nk} u_{nk}(r) \tag{2.8}$$

The $k$-independent term has for solution the energies of the bands $\varepsilon_{n0}$ computed at the $\Gamma$ point ($k = 0$):

$$H(k=0)u_{n0} = \varepsilon_{n0} u_{n0} \tag{2.9}$$

The operator $W(k) = \hbar^2 k^2/2m_0 + \hbar k \cdot p/m_0$ vanishes for $k = 0$ and commutes with the translation operator. The former property allows to write down a perturbation analysis for small $k$ values since the operator $W$ is adiabatically coupled with $H(0)$. The translational invariance $[W(k), T_T] = 0$ allows to write the perturbed solutions as a linear superposition of the unperturbed solutions:

$$u_{nk} = \sum_m c_{m}^{(n)}(k) u_{m0}. \tag{2.10}$$

In Eq.(2.7) it yields:

$$\sum_m \left( \frac{\varepsilon_{m0} - \varepsilon_{nk} + \hbar^2 k^2}{2m_0} \delta_{lm} + \frac{\hbar}{m_0} \langle m0|k \cdot p|m0 \rangle \right) c_{m}^{(n)}(k) = 0. \tag{2.11}$$

We have used that bra-ket notation for the matrix elements of the $k \cdot p$ operator. The latter reads in the position representation:

$$\langle 0|k \cdot p|m0 \rangle = \int_{\text{unit cell}} u_{n0}^*(r) k \cdot p u_{m0}(r). \tag{2.12}$$

where $\langle r|n k \rangle = u_{nk}(r)$.

From the expansion in Eq.(2.10) we have necessarily $c_{m}^{(n)}(0) = \delta_{nm}$. The non-diagonal matrix elements in Eq.(2.12) give rise to coherent superposition of zeroth order solutions.

In order to simplify the calculations we assume that the bands at $k = 0$ are non-degenerated: $\varepsilon_{n0} \neq \varepsilon_{m0}$ for $n, m = 1, 2, ...$.

Since the perturbation is adiabatically coupled to the system, the coefficients $c_{m}^{(n)}(k)$ are continuous functions of $k$. As $c_{m}^{(n)}(0) = \delta_{nm}$, the first non-trivial contribution to the
expansion of \( c_m^{(n)}(k) \) \((n \neq m)\) are linear: \( c_m^{(n)}(k) = \alpha_m \cdot k \). (we have dropped the band
index \((n)\) in \( \alpha_m \) in order to simplify the notation)

For the \( m = n \) term, we have \( c_m^{(n)}(0) = 1 \). For small values of \( k \), we can assume that
\( c_m^{(n)}(k) \approx 1 \). A linear term would give rise to higher order terms than \( k^2 \). We therefore set
\( \alpha_n = 0 \).

The first contributions to the perturbative expansion are found with the Ansatz:

\[
c_n^{(n)} = 1 \quad \text{and} \quad c_m^{(n)} = \alpha_m \cdot k, \quad \text{for } m \neq n.
\]

(2.13)

It is worth to mention that in Eq.(2.11) there is not diagonal term proportional to \( k \). The
first contributions to the perturbation expansion are therefore of order \( k^2 \). The \( k \cdot p \) matrix
element starts also with \( k^2 \) terms as \( c_m^{(n)} \) is linear in \( k \).

Before we evaluate the perturbation expansion we introduce a special notation. The latter
is found useful when the spin-orbit coupling is considered in the \( W(k) \) perturbation:

\[
\pi = p \quad \text{and} \quad \pi_{nm} = \langle n|\pi|m \rangle.
\]

(2.14)

We first evaluate Eq.(2.11) for \( l = n \). We insert the ansatz for \( c_m^{(n)}(k) \):

\[
\varepsilon_{nk} = \varepsilon_{n0} + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar}{m_0} \sum_{m \neq n} k \cdot \pi_{nm} \alpha_m \cdot k.
\]

(2.15)

We have restricted the sum to \( m \neq n \) as \( \alpha_n = 0 \). We now evaluate Eq.(2.11) for \( l \neq n \):

\[
0 = \left( \varepsilon_{l0} - \varepsilon_{nk} + \frac{\hbar^2 k^2}{2m_0} \right) c_l^{(n)}(k) + \frac{\hbar}{m_0} \left( \sum_{m \neq n} k \cdot \pi_{ln} c_n^{(n)}(k) + \sum_{m \neq n} k \cdot \pi_{lm} c_m^{(n)}(k) \right)
\]

\[
= \left( \varepsilon_{l0} - \varepsilon_{n0} + \mathcal{O}(k^2) \right) \alpha_l \cdot k + \frac{\hbar}{m_0} k \cdot \pi_{ln} + \frac{\hbar}{m_0} \sum_{m \neq n} k \cdot \pi_{lm} \alpha_m \cdot k
\]

\[
= \alpha_l \cdot k (\varepsilon_{l0} - \varepsilon_{n0}) + \frac{\hbar}{m_0} k \cdot \pi_{ln} + \mathcal{O}(k^2)
\]

(2.16)

This yields an expression for \( \alpha_l \cdot k \) up to the first order in \( k \):

\[
\alpha_l \cdot k = \frac{\hbar}{m_0} \varepsilon_{n0} - \varepsilon_{l0}.
\]

(2.17)

The perturbative expansion of the energy is obtained by inserting the above expression in
Eq.(2.15) with a correct remapping of the indices:

\[
\varepsilon_{nk} = \varepsilon_{n0} + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2}{m_0^2} \sum_{m \neq n} \frac{k \cdot \pi_{mn} \pi_{nm} \cdot k}{\varepsilon_{n0} - \varepsilon_{m0}}.
\]

(2.18)
This approximation is accurate only when the kinetic energy in the perturbed band \((n)\) is smaller than all gaps formed with the unperturbed bands \((m)\). This condition reads:

\[ |\varepsilon_{nk} - \varepsilon_n| \ll |\varepsilon_{n0} - \varepsilon_{m0}| \text{ for all } m = 1, 2, \ldots \]

It is important to remark that the \(k \cdot p\) perturbation of non-degenerated bands gives a parabolic energy dispersion in the vicinity of the \(\Gamma\)-point. We can rewrite Eq.(2.18) as:

\[ \varepsilon_{nk} = \varepsilon_{n0} + \frac{\hbar^2}{2} \sum_{\alpha\beta} k_\alpha \frac{1}{\mu_{\alpha\beta}} k_\beta. \]  

(2.19)

The indices \(\alpha\) and \(\beta\) are the cartesian coordinates \(\alpha, \beta = x, y, z\). The effective mass tensor \(\mu_{\alpha\beta}\) reads:

\[ \frac{1}{\mu_{\alpha\beta}} = \frac{1}{m_0} \delta_{\alpha\beta} + \frac{2}{m_0} \sum_{m \neq n} \pi_{mn\alpha} \pi_{mn\beta}. \]  

(2.20)

where \(\pi_{mn\alpha} = \hat{k}_\alpha \cdot \langle n0|\pi|m0\rangle\), the unit vector \(\hat{k}_\alpha\) pointing in the \(\alpha\)-direction.

In Fig.(2.5) the \(k \cdot p\) approximation is illustrated on the band diagram of GaAs.

In III-V bulk semiconductors the \(k \cdot p\) approximation is performed at the \(\Gamma\)-point as the gap is direct. However the method can be used for indirect-gap materials, the perturbative expansion being made at an other high-symmetry point, corresponding to the minimum of the lowest-lying conduction band. The calculations will give different values for the effective mass, but the approximation framework is still valid.

In III-V semiconductors, the spin-orbit coupling cannot be neglected. This yields an additional term to \(H(k=0)\) and to the perturbation part \(W(k)\) in the hamiltonian. The modified operators read:

\[ H(k=0) = \frac{\hbar^2 k^2}{2m_0} + V(r) + \frac{\hbar}{4m_0 c^2} (\sigma \wedge \nabla V) \cdot p \]  

(2.21)

\[ W(k) = \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar k}{m_0} \cdot \left( p + \frac{\hbar}{4m_0 c^2} \sigma \wedge \nabla V \right) \]

Where \(\sigma\) is the spin operator (Pauli matrices). The energy of the band-edges \((k=0)\) are modified. Depending on the value of the total (orbital + spin) angular momentum some degeneracies are lifted.

The perturbation \(W(k)\) has still the form \(k \cdot \ldots\). We therefore obtain the same type of quadratic energy dispersion as in Eq.(2.19) but with new band-edge energies \(\varepsilon_{n0}^{\text{spin-orbit}}\) and a modified \(\pi\) vector:

\[ \pi = p + \frac{\hbar}{4m_0 c^2} \sigma \wedge \nabla V \]  

(2.22)
Chapter 2. Band structure of bulk III-V semiconductors

Figure 2.5: The band diagram of GaAs is shown. The shaded area in orange represents the filled valence bands. The gap is shown in grey. The conduction band is highlighted in blue. At the Γ point, the energy dispersion approximated with the $\mathbf{k} \cdot \mathbf{p}$ model is illustrated is red. The basic figure was taken from [80].

2.5 The Kane model and the conduction-band non-parabolicity

It is clear by looking at Eq.(2.19) or simply at Fig.(2.5) that more the effective mass is heavier, more the kinetic energy varies slowly with $k$. The $\mathbf{k} \cdot \mathbf{p}$ approximation is therefore valid for a larger range of $k$ values. For lighter effective masses, the kinetic term $\varepsilon_{nk} - \varepsilon_{n0}$ becomes faster comparable to the band gaps at the Γ point.

In this case the influence of the other bands (the valence bands which are the closest to the conduction band) on the energy dispersion in the conduction band becomes rapidly more
important. It is tempting to go beyond the quadratic approximation in the perturbative expansion but this yields tedious calculations.

An elegant method that accounts for conduction-band non-parabolicity was proposed by E. Kane [54]. The latter gives accurate results without having to perform complicated higher order perturbative expansions. In InSb the topmost valence bands and the conduction band are very close and well-separated from other bands. The perturbation $W(k)$ can be diagonalized exactly for this subset of bands. The coupling to the other (remote) bands is treated as a usual second-order perturbation.

We consider the hamiltonian with the spin-orbit coupling in Eq.(2.21). As previously we assume that we know the band-edge energies $\varepsilon_{n0}$ and the corresponding periodic wavefunctions $u_{n0}$ that diagonalize the crystal hamiltonian (without spin-orbit coupling):

$$H_{\text{crystal}} U_i = E_i U_i \quad \text{with} \quad H_{\text{crystal}} = \frac{p^2}{2m_0} + V(r)$$

We assume that a subset of band-edges are well-separated from the others. This subset is constituted by $s$-type wavefunctions for the conduction band and $p$-type wavefunctions for the closest valence bands. The $s/p$-type refers to the eigen-wavefunctions of the orbital momentum. In Eq.(2.23) the conduction band states are singly degenerated (double degeneracy with spin) and are written $|S \uparrow\rangle$ and $|S \downarrow\rangle$. The valence states have a triple degeneracy (sixfold with spin) and are written $|X \uparrow\rangle$, $|Y \uparrow\rangle$, $|Z \uparrow\rangle$ and $|X \downarrow\rangle$, $|Y \downarrow\rangle$, $|Z \downarrow\rangle$. For commodity the energy of the conduction states is set to zero: $H_{\text{crystal}}|S \uparrow\rangle = E_s|S \uparrow\rangle$ with $E_s = 0$. We assume that the valence states have an energy $-\varepsilon_0$: $H_{\text{crystal}}|X_i \uparrow\rangle = -\varepsilon_0|X_i \uparrow\rangle$ for $X_i = X, Y, Z$. The letter $X, Y, Z$ are used to underline the directional character of the $p$-type wavefunctions in the direction $x, y$ and $z$.

The degeneracy of the subspaces associated with the eigenvalues $E_s$ or $E_p$ relies on a group-theory argument [98] (tetrahedral symmetry corresponding to point group $T_d$).

The hamiltonian in Eq.(2.21), can be written:

$$H = \frac{p^2}{2m_0} + V(r) + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar k \cdot p}{m_0} + \frac{\hbar}{4m_0 c^2} (\sigma \wedge \nabla V) \cdot p + \frac{\hbar^2}{4m_0^2 c^2} k \cdot (\sigma \wedge \nabla V)$$

The two first terms give the crystal hamiltonian $H_{\text{crystal}}$, the third is the free-electron (heavy particles) dispersion term, the fourth term is the $k \cdot p$ coupling, the fifth is the atomic spin-orbit coupling that depends on the atomic momentum $p$ and the sixth is the crystal spin-orbit coupling.

In the Kane model we diagonalize exactly the following part of the hamiltonian for the
subset of bands defined above:

\[
H_{\text{Kane}} = \frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}) + \frac{\hbar^2 \mathbf{k}^2}{2m_0} + \frac{\hbar}{4m_0 c^2} (\sigma \wedge \nabla V) \cdot \mathbf{p} + \frac{\hbar \mathbf{k} \cdot \mathbf{p}}{m_0}
\]  

(2.25)

The \( \mathbf{k} \)-dependent spin-orbit coupling is known to give a linear dependence on \( \mathbf{k} \) in the valence bands energy dispersions. We neglect this contribution.

Before we evaluate the matrix elements of \( H_{\text{Kane}} \) for the conduction- and valence-band states, it is convenient to rewrite them as combinations of eigenfunctions \( |JM\rangle \) of the total angular momentum operator \( \mathbf{J} = \mathbf{L} + \mathbf{S} \). The spin-orbit coupling term is therefore diagonal. Practically it is a unitary transformation (basis change) is the sub-space associated with eigen-energies \( E_s \) and \( E_p \).

For \( s \)-states, the orbital quantum number is \( l = 0 \) and the spin quantum number is \( s = \frac{1}{2} \). We therefore only have one value of \( J = \frac{1}{2} \), with \( M = -\frac{1}{2}, \frac{1}{2} \). With the special convention in usage in [8] that follows [54], we have:

\[
| \frac{1}{2}, \frac{1}{2} \rangle_s = i | S \uparrow \rangle \quad | \frac{1}{2}, -\frac{1}{2} \rangle_s = i | S \downarrow \rangle
\]  

(2.26)

We have added a \( s \) subscript in order to avoid confusion with the \( p \)-states.

For \( p \)-states, the orbital number is \( l = 1 \) and the spin number is \( s = \frac{1}{2} \). The quantum number \( J \) can therefore take two values: \( J = \frac{1}{2} \) (doublet) and \( J = \frac{3}{2} \) (quadruplet). In order to express the \( |X_i\rangle \) in terms of spherical harmonics we have to remark that when the \( z \)-axis is the quantization axis, we have:

\[
| l = 1, m_l = 1 \rangle \propto \frac{1}{\sqrt{2}} (|X\rangle + i|Y\rangle) \quad | l = 1, m_l = 0 \rangle \propto |Z\rangle \quad | l = 1, m_l = -1 \rangle \propto \frac{1}{\sqrt{2}} (|X\rangle - i|Y\rangle)
\]  

(2.27)

We have kept the proportionality symbol although the normalization factor can be absorbed in the definition of the \( |X_i\rangle \).

The total momentum eigen-states can be expressed with the \( |X_i \uparrow, \downarrow\rangle \). For \( J = \frac{3}{2} \) and...
$J = \frac{1}{2}$ we have:

\begin{align*}
|\frac{3}{2}, +\frac{3}{2}\rangle_p &= \frac{1}{\sqrt{2}}|(X + iY) \uparrow\rangle \\
|\frac{3}{2}, +\frac{1}{2}\rangle_p &= -\sqrt{\frac{2}{3}}|Z \uparrow\rangle + \frac{1}{\sqrt{6}}|(X + iY) \downarrow\rangle \\
|\frac{3}{2}, -\frac{1}{2}\rangle_p &= -\frac{1}{\sqrt{6}}|(X - iY) \uparrow\rangle - \sqrt{\frac{2}{3}}|Z \downarrow\rangle \\
|\frac{3}{2}, -\frac{3}{2}\rangle_p &= \frac{1}{\sqrt{2}}|(X - iY) \downarrow\rangle \\
|\frac{1}{2}, +\frac{1}{2}\rangle_p &= \frac{1}{\sqrt{3}}|(X + iY) \downarrow\rangle + \frac{1}{\sqrt{3}}|Z \uparrow\rangle \\
|\frac{1}{2}, -\frac{1}{2}\rangle_p &= -\frac{1}{\sqrt{3}}|(X - iY) \uparrow\rangle + \frac{1}{\sqrt{3}}|Z \downarrow\rangle.
\end{align*}

(2.28)

The normalization is not the one used in [20]. However, we have preferred to use the same wavefunctions than [8, 54] in order to keep the consistency with the literature.

Before we compute the matrix elements of the Kane hamiltonian we choose a sorting for the basis. We sort the $s$ and $p$ eigenstates with positive and negative $M_J$ values. This choice will benefit since if we choose the quantization axis $z$ parallel to the crystal momentum $k$, the Kane matrix will become block-diagonal and the system can be reduced to a $4 \times 4$ hamiltonian with spin degeneracy. We therefore choose the basis:

\begin{align*}
\left\{ |\frac{1}{2}, \frac{1}{2}\rangle_s, |\frac{3}{2}, \frac{1}{2}\rangle_p, |\frac{3}{2}, \frac{3}{2}\rangle_p, |\frac{1}{2}, \frac{1}{2}\rangle_p, |\frac{1}{2}, -\frac{1}{2}\rangle_s, |\frac{3}{2}, -\frac{1}{2}\rangle_p, |\frac{3}{2}, -\frac{3}{2}\rangle_p, |\frac{1}{2}, -\frac{1}{2}\rangle_p \right\}
\end{align*}

(2.29)

We first consider the Kane hamiltonian for $k = 0$. As the atomic spin-orbit coupling is diagonal in the $|JM\rangle$ basis the spin-orbit interaction will only lift the degeneracy.

The splitting energy is given by the evaluation of the spin-orbit term: $\frac{\hbar}{4m_e c^2} (\sigma \wedge \nabla V) \cdot p$. This contribution is zero for $s$-states and the conduction band states therefore still have a zero energy. This contribution is also zero for the quadruplet of $p$-states ($J = \frac{3}{2}$) that keep an energy of $-\varepsilon_0$. However the $p$-states doublet ($J = \frac{1}{2}$) is lowered by the interaction to $-\varepsilon_0 - \Delta$. The splitting energy being given by:

\begin{align*}
\Delta = \frac{3\hbar}{4m_e c^2} \left\langle X \left( \frac{\partial V}{\partial x} p_y - \frac{\partial V}{\partial y} p_x \right) Y \right\rangle.
\end{align*}

(2.30)

The calculations are not reproduced here. The above expressed was taken from the original paper of Kane [54].

We now consider the Kane hamiltonian for non-vanishing $k$. The free-electron dispersion is diagonal but the $k \cdot p$ couples the conduction and the valence states. We omit the diagonal
free-electron dispersion $\frac{\hbar^2 k^2}{2m_0}$ for commodity. The $8 \times 8$ matrix of the Kane hamiltonian is given by:

$$
\begin{pmatrix}
0 & -\sqrt{\frac{2}{3}} P \hbar k_x & \frac{1}{\sqrt{3}} P \hbar k_y & 0 & -\frac{1}{\sqrt{3}} P \hbar k_z & 0 & -\sqrt{\frac{2}{3}} P \hbar k_-
-\sqrt{\frac{2}{3}} P \hbar k_x & -\varepsilon_0 & 0 & 0 & \frac{1}{\sqrt{3}} P \hbar k_+ & 0 & 0 & 0
\frac{1}{\sqrt{3}} P \hbar k_y & 0 & 0 & -\varepsilon_0 - \Delta & \sqrt{\frac{2}{3}} P \hbar k_+ & 0 & 0 & 0
0 & \frac{1}{\sqrt{3}} P \hbar k_+ & 0 & \sqrt{\frac{2}{3}} P \hbar k_+ & 0 & -\sqrt{\frac{2}{3}} P \hbar k_z & P \hbar k_+ & \frac{1}{\sqrt{3}} P \hbar k_z
-\frac{1}{\sqrt{3}} P \hbar k_+ & 0 & 0 & 0 & -\sqrt{\frac{2}{3}} P \hbar k_z & -\varepsilon_0 & 0 & 0
0 & 0 & 0 & 0 & \sqrt{\frac{2}{3}} P \hbar k_+ & -\varepsilon_0 & 0 & 0
-\sqrt{\frac{2}{3}} P \hbar k_+ & 0 & 0 & 0 & \frac{1}{\sqrt{3}} P \hbar k_z & 0 & 0 & -\varepsilon_0 - \Delta
\end{pmatrix}
$$

(2.31)

We have the following definitions:

$$
k_\pm = \frac{1}{\sqrt{2}} (k_x \pm ik_y)
$$

$$
P = -\frac{i}{m_0} \langle S|p_x|X \rangle = -\frac{i}{m_0} \langle S|p_y|Y \rangle = -\frac{i}{m_0} \langle S|p_z|Z \rangle.
$$

(2.32)

In order to find the eigen-energies from the secular equation of the above matrix, we define the energy $\lambda(k) = \varepsilon(k) - \frac{\hbar^2 k^2}{2m_0}$. The eigen-solutions are determined by:

$$
\lambda(k) = -\varepsilon_0
$$

(2.33)

$$
\lambda(k) \left( \lambda(k) + \varepsilon_0 \right) \left( \lambda(k) + \varepsilon_0 + \Delta \right) = \hbar^2 k^2 P^2 \left( \lambda(k) + \varepsilon_0 + \frac{2\Delta}{3} \right).
$$

(2.34)

As we have noticed before, the dispersion relation depends on $k^2$ only. The choice of the quantization axis along the crystal momentum $k$ makes the Kane matrix $4 \times 4$ block-diagonal as $k_\pm$ vanishes and $k_z = k$. This choice that may appear arbitrary in bulk material will be very efficient when we will consider planar heterostructures grown in the same direction ($z$-direction). In heterostructures an arbitrary quantization axis will yield complicated results.

Approximated solutions [8] can be derived for $\lambda(k)$. In particular the effective masses when $k \to 0$. We will not focus on these solutions as some numerical examples are given in next paragraph (2.5.1). The important consequence is that the energy dispersion of the conduction band obtained in the Kane model is no more parabolic. This effect can be reformulated by an energy dependence of the mass:

$$
E = \frac{\hbar^2 k^2}{2m(E)}
$$

(2.35)
2.5. The Kane model and the conduction-band non-parabolicity

This result is a crucial improvement of the model since it cannot be obtained by a finite order perturbation series. It has also a large impact on the band structure as the deviation from a parabolic model is significant for III-V alloys.

2.5.1 Numerical results for various III-V alloys

The matrix element of the velocity operator $P/m_0$ is usually defined with the Kane’s energy $E_p$. The definition of the latter varies in literature. Here we choose the convention of [89]. If the matrix element coupling the conduction and the valence band is defined by $p_{cv} = m_0 P$, then:

$$p_{cv} = i \sqrt{\frac{m_0 E_p}{2}}$$  \hspace{1cm} (2.36)

The numerical evaluation of the energy dispersion $\varepsilon(k)$ in Eq.(2.33-2.34) are shown in Fig.(2.6) for various materials. The band parameters are given in Tab.(2.1).

![Figure 2.6: The solution of the Kane matrix with J collinear to k is shown for various III-V alloys: InSb, InAs, InP and GaAs. The computation is made in the vicinity of the Γ-point. The gap energy is reported as $\varepsilon_0$ and the spin-orbit splitting as Δ. The conduction band is shown in green with index (C). The valence bands are: heavy hole (HH), light hole (LH) and split-off (SO). The non-parabolicity of the conduction band is very clear for narrow ago materials: InSb and InAs.](image-url)
Chapter 2. Band structure of bulk III-V semiconductors

<table>
<thead>
<tr>
<th></th>
<th>( \varepsilon_0 ) (eV)</th>
<th>( E_p ) (eV)</th>
<th>( \Delta ) (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>InSb</td>
<td>0.235</td>
<td>22.49</td>
<td>0.81</td>
</tr>
<tr>
<td>InAs</td>
<td>0.418</td>
<td>21.11</td>
<td>0.38</td>
</tr>
<tr>
<td>InP</td>
<td>1.424</td>
<td>17</td>
<td>0.11</td>
</tr>
<tr>
<td>GaAs</td>
<td>1.519</td>
<td>22.71</td>
<td>0.341</td>
</tr>
</tbody>
</table>

Table 2.1: Kane model parameters for various III-V alloys. The values were taken from [107]

The Kane parameters are usually not computed from basic quantum models but are deduced from optical measurements. For \( k = 0 \) the energies of the band edges are given by: \( E_C = 0, E_{HH} = -\varepsilon_0, E_{LH} = -\varepsilon_0 \) and \( E_{SO} = -\varepsilon - \Delta \).

For narrow gap materials, or for high kinetic energies, the conduction band dispersion is clearly no more parabolic. This is explicitly due to the \( \mathbf{k} \cdot \mathbf{p} \) coupling with the valence bands. Although it may be interesting to keep the general solution of the Kane model, we will reduce the problem to an effective two-band model. The latter is found accurate enough in this work.

2.6 Three-band model: conduction, light-hole and split-off bands

If we look to the Kane hamiltonian matrix in Eq.(2.31) and to Fig.(2.6) we see that the heavy-hole (HH) band is decoupled from the other bands. The HH band is expected to have a weak impact on the dynamics of the systems studied in this work that are based on conduction electrons (light particles). Even if the coupling to the remote bands will reduce the mass of the HH band, the latter is still very large compared to the other effective masses in the Kane model.

We therefore only consider the conduction (C), light-hole (LH) and split-off (SO) bands. The reduced matrix (without the spin) reads [89]:

\[
H_{C,LH,SO} = \begin{pmatrix}
E_c & \sqrt{\frac{2}{3} \frac{\hbar}{m_0}} p_{cv} k & -\frac{1}{\sqrt{3} \frac{\hbar}{m_0}} p_{cv} k \\
-\sqrt{\frac{2}{3} \frac{\hbar}{m_0}} p_{cv} k & E_{lh} & 0 \\
\frac{1}{\sqrt{3} \frac{\hbar}{m_0}} p_{cv} k & 0 & E_{so}
\end{pmatrix}
\] (2.37)

We have used an explicit notation for the band edges \( E_c = 0, E_{lh} = -\varepsilon_0 \) and \( E_{so} = -\varepsilon - \Delta \). The notation \( p_{cv} = m_0 P \) was introduced previously. We have further simplified the model
by neglecting the free electron dispersion $\frac{h^2k^2}{2m_0}$. The latter has a very weak variation as shown by the HH band in Fig.(2.6). Its contribution in terms of energy is only \[89\] of order $(E_c - E_{lh, so})/E_p \ll 1$.

We can however model accurately the non-parabolicity of the conduction band. We consider $H_{C,LH,SO} \psi = E \psi$, with $E$ being an eigenvalue and $\psi = (\phi_c, \phi_{lh}, \psi_{so})$ the associated eigen-wavefunctions (corresponding to eigen-states $|JM\rangle$ of the total angular momentum). The latter being the periodic function $u$ in the Bloch wave. We can express the light-hole and split-off parts in terms of the conduction part:

$$
\phi_{lh} = -\sqrt{\frac{2}{3}} \frac{1}{E - E_{lh}} \frac{h}{m_0} p_{cv} k \phi_c
$$

$$
\phi_{so} = \frac{1}{\sqrt{3}} \frac{1}{E - E_{so}} \frac{h}{m_0} p_{cv} k \phi_c.
$$

Substituted back into the equation for $\phi_c$ we obtain the energy dispersion for the electrons in the conduction band:

$$
E = E_c + \frac{h^2k^2}{2m_0} \left( \frac{2}{3} \frac{E_p}{E - E_{lh}} + \frac{1}{3} \frac{E_p}{E - E_{so}} \right)
$$

$$
= E_c + \frac{h^2k^2}{2m(E)}.
$$

where we have used $E_p = \frac{2}{m_0} p_{cv}^2$. The energy-dependent effective mass reads:

$$
m^{-1}(E) = \frac{1}{m_0} \left( \frac{2}{3} \frac{E_p}{E - E_{lh}} + \frac{1}{3} \frac{E_p}{E - E_{so}} \right).
$$

For $k \to 0$, we obtain:

$$
m^{-1}(E_c) = \frac{1}{m_0} \left( \frac{2}{3} \frac{E_p}{E_c - E_{lh}} + \frac{1}{3} \frac{E_p}{E_c - E_{so}} \right)
$$

$$
= \frac{1}{m_0} \frac{E_p}{3} \left( \frac{2}{\varepsilon_0} + \frac{1}{\varepsilon_0 + \Delta} \right),
$$

where the usual notation of the Kane model have been restored. The fraction $m(E)/m(E_c)$ can be expressed in terms of the excess energy $E'$ in the conduction band $E_c(E') = E_c + E'$:

$$
\frac{m(E')}{m(E_c)} = \left( \frac{2}{\varepsilon_0} + \frac{1}{\varepsilon_0 + \Delta} \right) \left( \frac{2}{E' + \varepsilon_0} + \frac{1}{E' + \varepsilon_0 + \Delta} \right)^{-1}.
$$

The ratio does not depend on $E_p$. This result is plotted for various III-V alloy in Fig.(2.7).
Figure 2.7: The energy-dependent mass $m(E)$ is plotted for various III-V alloys: InAs, InSb and GaAs. For each alloy the mass is computed in the three-band model (conduction, light-hole, split-off bands) in solid line and in the effective two-band model (conduction, upper effective valence bands) in dashed line. The shaded areas show for the discrepancy between the models. In both models the effective mass has a linear dependence. The effective two-band model has a good accuracy for InAs and GaAs. For InSb the mass is not accurately predicted by the two-band model. The accuracy is connected with error estimations in Tab(2.2).

The numerical results are interesting since the mass has a strong dependence on $E'$. It doubles for an energy of 0.52 eV typical to $\text{Ga}_{0.47}\text{In}_{0.53}\text{As} / \text{Al}_{0.48}\text{In}_{0.52}\text{As}$ heterostructures. As expected, the coupling with the valence bands gives quantitatively very significant results.

Another property is the linear character of the mass with respect to the excess energy $E'$. The energy-dependent mass can therefore be modeled empirically by a linear relation for most of the III-V alloys. We write: $m(E') = m(0) (1 + \alpha E')$, with $\alpha$ being a constant for a given alloy. The constant is obtained by a Taylor expansion of Eq.(2.42) as a function
of $E'$. We have for the linear term in $E'$:

$$\alpha = \frac{1}{\varepsilon_0} \frac{1 + \frac{1}{2} \delta^2}{1 + \frac{1}{2} \delta} \approx \frac{1}{\varepsilon_0} \quad \text{therefore} \quad m(E') \approx m(0) \left( 1 + \frac{E'}{\varepsilon_0} \right)$$

(2.43)

where $\delta = \frac{\epsilon_0}{\varepsilon_0 + \Delta}$.

The non-parabolicity coefficient $\gamma$ is usually introduced. It is defined by considering the energy dispersion as a function of $k$. If we set the conduction-band edge energy to zero: $E_c = 0$, we can write:

$$E(k) = \frac{\hbar^2 k^2}{2m(E(k))} \approx \frac{\hbar^2 k^2}{2m(0)} \left( 1 - \frac{E(k)}{\varepsilon_0} \right) = \frac{\hbar^2 k^2}{2m(0)} \left( 1 - \gamma k^2 \right).$$

(2.44)

We have assumed that $E/\varepsilon_0 \ll 1$. The coefficient $\gamma$ is therefore defined by the coefficient of the quartic term $k^4$. It reads explicitly:

$$\gamma = \frac{\hbar^2}{2m(0)\varepsilon_0}$$

(2.45)

This parameter is very practical in the calculations of the band-structure, as it links fundamental parameters between them.

## 2.7 Effective two-band model

As the three-band model give accurate results, it is tempting to do more simplifications. The cruder one would be to neglect the split-off component. This is a bad approximation. It is more fruitful to consider instead an effective two-band model [89]. We consider the following unitary transformations between the light-hole and split-off components:

$$\phi_v = \sqrt{\frac{2}{3}} \phi_{lh} - \sqrt{\frac{1}{3}} \phi_{so}$$

$$\phi_{v'} = \sqrt{\frac{1}{3}} \phi_{lh} + \sqrt{\frac{2}{3}} \phi_{so}.$$  

(2.46)

This allows to re-express the three-band hamiltonian in Eq.(2.37) in terms of the new valence states $v$ and $v'$. We have:

$$H_{C,V,V'} = \begin{pmatrix} E_c & \frac{\hbar}{m_0} p_{cv} k & 0 \\ -\frac{\hbar}{m_0} p_{cv} k & E_v & \eta \\ 0 & \eta & E_{v'} \end{pmatrix}$$

(2.47)
Chapter 2. Band structure of bulk III-V semiconductors

<table>
<thead>
<tr>
<th></th>
<th>InAs</th>
<th>InSb</th>
<th>GaAs</th>
<th>InP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error [%]</td>
<td>9</td>
<td>37.3</td>
<td>0.9</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table 2.2: The error in the effective two-band model is computed for various III-V alloys. In terms of the Kane Hamiltonian energies the error reads: $2\Delta^2(3\varepsilon_0 + 2\Delta)^{-1}(3\varepsilon_0 + \Delta)^{-1}$.

With the definitions:

$$E_v = \frac{2E_{lh} + E_{so}}{3}, \quad E_{v'} = \frac{E_{lh} + 2E_{so}}{3}, \quad \eta = \frac{\sqrt{2}(E_{lh} - E_{so})}{3}. \quad (2.48)$$

After this unitary transformation the conduction band is coupled by a $k \cdot p$ term to the effective valence band $v$. The latter being coupled to $v'$ with matrix elements $\eta$. The nice effect of this transformation reside in that the contribution of $|\phi_{v'}|$ to the energy is given by the ratio $\eta^2/(E_c - E_v)(E_c - E_{v'})$. The same correction is found for the normalization of the total wavefunction. In Tab.(2.2) the deviation from the three-band model is reported.

For main III-V alloys this error is negligible and the effective two-band model will give accurate results. However for InSb this error is large (about 37 %) and the approximation will not give accurate results.

The effective two-band model is obtained by neglecting the band $v'$. The Hamiltonian is reduced to:

$$H_{C,V} = \left( \begin{array}{c} E_c \\ -\frac{h}{m_0} p_{cv} k \\ E_v \end{array} \right). \quad (2.49)$$

The energy-dependent effective mass for the conduction-band can be obtained by looking to the eigen-problem $H_{C,V}\tilde{\psi} = E\tilde{\psi}$; where $\tilde{\psi} = (\phi_c, \phi_v)$. We get:

$$E_c(k) = E_c + \frac{\hbar^2 k^2}{2m(E)} \quad \text{where} \quad m(E) = m_0 \frac{E - E_v}{E_p}, \quad E_v = \frac{2E_{lh} + E_{so}}{3}. \quad (2.50)$$

From the three-band model we know that the mass have a linear dependence with the excess energy $E'$ in the conduction band. We have $m(E') = m(0)(1 + \alpha E')$. We define the effective gap $\bar{E}_g = E_c - E_v$ between the conduction-band edge and the effective valence-band edge. With this definition and a Taylor expansion of $m(E')/m(0)$ up to the first order in $E'$, we find:

$$m(E') = m(0) \left(1 + \frac{E'}{\bar{E}_g} \right) \quad \text{with} \quad m(0) = m_0 \frac{\bar{E}_g}{E_p}. \quad (2.51)$$

For the effective two-band model this relation is exact. The mass $m(0)$ for $k \to 0$ can be computed, but we usually prefer to use the measured values. The measured value of the mass at the edge of the conduction band is intrinsically renormalized by all interactions.
inside the crystal. The model is here only used in order to obtain the functional dependence of mass with respect to the excess energy. This mixed approach gives accurate numerical results.

A more pragmatic approach should involve two measurements of the mass at different energies. Or at least two measurements that are connected to the mass (energy differences between two conduction subbands in a heterostructure). As the dependence is linear, the slope can be obtained as a fitting parameter. In this way, the uncertainty of the model is partially removed.

As the mass of the three-band model and the effective two-band model are both linear, we can define a non-parabolicity coefficient $\bar{\gamma}$ for the effective gap $\bar{E}_g$.

If we know the non-parabolicity coefficient $\bar{\gamma}$ and the mass $m(0)$, we can deduce the effective gap and the Kane energy with the relations:

$$\bar{E}_g = \frac{\hbar^2}{2m(0)\bar{\gamma}} \quad \text{and} \quad E_p = \frac{m_0}{m(0)}\bar{E}_g.$$  \hspace{1cm} (2.52)

### 2.8 Ternary and quaternary III-V alloys: the virtual crystal approximation

Until now we have considered crystalline bulk III-V semiconductors, like InAs, GaAs, InSb, InP, etc. Nevertheless most of the heterostructures based on III-V materials will involve ternary $\text{AB}_1\text{C}_x$ of even quaternary $\text{AB}_x\text{CD}_y$ alloys, where $x + y + z = 1$. From a microscopic point of view these solid solutions are not crystalline as if there is always an atom of type A in the elementary cell, the other atom can be of type B, C or D. We assume that the distribution of B, C and D atoms is random but gives in average the fraction $x$, $y$ and $z$. The crystal potential $V(\mathbf{r})$ loses the translational invariance of the Bravais lattice. In this perspective, the electronic states cannot be computed by using the Bloch theorem and energy bands or any crystal-related properties cannot be defined.

It is more fruitful to understand the alloy in terms of a mean-field quantity. The latter was already used implicitly when the existence of a one-electron crystal potential $V(\mathbf{r})$ was assumed in the Schrödinger stationary equation in Eq.(2.2). The many-body problem regarding the bonding of molecular orbitals, the effect of screening of the nuclear potentials by the sea of inner and outer electrons, was expressed as an effective one-electron (a crystal quasi-particle) problem. For a ternary or higher order alloy we can consider that an electron of the crystal sees an averaged potential given by the respective crystal potential
of each constituent of the alloy. For a ternary alloy we can write:

\[ \langle V \rangle = V_A + xV_B + (1 - x)V_C. \]  

(2.53)

The potential of the alloy is an averaged potential given by a linear interpolation between each potential of its constituents. This approximation is called the virtual crystal approximation. In this approximation, the crystal potential recovers its translational invariance and the Bloch theorem can be used to define a band-structure. The microscopic random potential have of course consequences: the band-structure can be blurred and the Bloch waves are expected to be scattered by the compositional variation of the crystal (5.3.2).

The basic constituents AB and AC have a close lattice period and orbital structure, fundamental quantities like the fundamental gap energy can even be approximated as a linear interpolation of the basic gaps:

\[ E_{g_{AB,C1-x}}(x) = xE_{g_{AB}} + (1 - x)E_{g_{AC}}. \]  

(2.54)

Although a linear rule may be accurate for some alloys like the famous Hg_{1-x}Cd_xTe, the actual value of the gap, or other relevant quantities, are tabulated for each fraction \( x \). One expects that the \( E_g(x) \) will be strongly non-linear for mismatched lattice constants as in InAs_{1-x}Sb_x. But even if the lattice constants are very close like in the Al_xGa_{1-x}As, the fundamental gap of the GaAs is found at the \( \Gamma \) point, while the fundamental gap of the AlAs is found at the X point. As a consequence of the mismatched band structure, the resulting fundamental gap of the alloy Al_xGa_{1-x}As is non-linear in the alloy fraction \( x \).

In this work we will be mainly concerned with Ga_{0.47}In_{0.53}As and Al_{0.48}In_{0.52}As. The fractions of these alloys are explicitly chosen so that their lattice constants are matched on the InP lattice constant. These alloys are mainly used in mid-infrared quantum cascade lasers.

In Fig.(2.8) the alloy phase diagram for various semiconductor quaternary alloys is shown.
2.8. Ternary and quaternary III-V alloys: the virtual crystal approximation

Figure 2.8: This phase diagram shows the fundamental energy gap versus the lattice constant for various semiconductor alloys. The particular case of AlInAs and GaInAs lattice matched on InP is explicitly shown. The wavelength corresponding to the fundamental gap is shown on the right axis [97].
Chapter 3

Heterostructures and the envelop function approximation

In this part we introduce briefly the envelop function formalism. The latter allows to describe the quantum states in a heterostructure. We first define what is a heterostructure. The simplest definition is the spatial composition of two bulk materials. Here we assume that a bulk semiconductor $A$ fills the space for $z < z_0$ and that a second bulk semiconductor $B$ fills the space for $z > z_0$. The plane defined by $z = z_0$ is the interface between the two materials. As in the case of the Bravais lattice and the crystalline structure, we need to complete this mathematical definition by looking microscopically to the interface between the materials. For simplicity we assume that $A$ and $B$ materials have the same lattice constant. The situation is illustrated in Fig. (3.1).

At the interface we expect that the atoms of both semiconductors will form hybrid chemical bonds. The latter enable the formation of a continuous solid-solution. The quality of the interface tightly depends on the crystal growth conditions. The growth itself is a very large and subtle subject and we will not give a lot of details here. Nevertheless, in the MBE (Molecular Beam Epitaxy) growth, the growth speed (number of atomic monolayer deposited in a second), the temperature of the substrate and the chemical type (atoms, alloy fraction) of the crystal are crucial parameters that determine both the bulk and the interface quality, in particular the sharpness of the latter. At microscopic scales the size of the interface, defined by the spatial region where the atomic potential changes from $V_A$ to $V_B$, cannot be reduced below one monolayer: the lattice constant of the grown materials. However in the virtual crystal approximation where the continuity of the solid-solution is assumed, the averaging allows to define compositional variations below one mono-layer of bulk material. An infinitesimally sharp interface cannot therefore exist in a real system. It is worth to mention that the chemical composition is not constant in a few monolayers
Figure 3.1: A heterostructure between two semiconductors A (with composition QP) and B (with composition MP) is shown. The interface between the two materials is mathematically the plane located at $z = z_0$. The lattice constant is the same of A and B materials. Hybrid bonds at the interface are given with bonding of atoms Q-P-M. The conduction-band edge is shown as a function of $z$. A discontinuity $V_b$ occurs at the interface as the conduction bands do not line up. This illustration is inspired from [8].

around the interface plane. The interface has a roughness related to the growth conditions. The interface plane is not defined precisely at $z = z_0$ and fluctuate over a few angstroms around $z_0$. The interface is usually modeled by randomly distributed islets of A or B materials where the planar radius of the islets has a gaussian correlation and an averaged height in the z-direction (5.3.1). Nevertheless for a particular sample a continuous crystal potential $V(x, y, z)$ can be defined for the whole heterostructure. The interface position can therefore be defined by contours of $V$ defined by: $V_0 = V(x, y, z_0(x, y))$. The interface position being defined by $z = z_0(x, y)$.

The modeling of the interface can be very complicated and depends on the physics of the growth. An important question is whether accurate results can be obtained by: treating formally mathematical interfaces at average positions, computing the heterostructure states and afterwards modeling the roughness as a perturbation that broadens optical transitions and cause scattering between quantum states. This question is open for a
large range of heterostructures. Nowadays however the accuracy of the growth enables this perturbative modeling in major binary systems in which quantum cascade lasers are fabricated.

In the heterostructure in Fig.(3.1), the misalignment of conduction-band edges between materials A and B causes a discontinuity $-V_b$ in the conduction-band profile in the $z$-direction. If we consider an electron with an energy $E < V_b$, it is found in an allowed region of the band-structure in material B, while it is found in a forbidden region in material A. The alternation of materials acting as a barrier (A) referred from of well-acting material (B) allows to build quantum wells in the conduction band profile. The latter confine electrons in a spatial region of the heterostructure. The aim of this section is to develop formally the equations of motion of an electron in an arbitrary heterostructure.

We will consider the envelop function approximation. While the latter is very intuitive and allows to design a heterostructure for a particular purpose (like quantum cascade lasers), it is limited to band-structure valleys at high-symmetry points of the Brillouin zone. This model is accurate in the description of heterostructure quantum states in the vicinity of a high symmetry point ($\Gamma$, $X$, $L$, ...) but fails to model coupling of heterostructure states belonging to different valley, in particular when degeneracy occurs. This issue is addressed by computing the band-structure of the heterostructure on a larger portion of the first Brillouin zone. Microscopical tight-binding or pseudopotential models are adapted for this purpose, although they are less intuitive.

![Diagram](image.png)

Figure 3.2: Type I and type II are shown for a simple binary heterostructure. (a) In the type I, the ABA structure is a quantum well for both the conduction electrons and the valence holes. (b) In type II, the ABA is a well for the conduction-band electrons, while it is a barrier for valence-band holes.
We can distinguish two types of heterostructures: type I and type II, as shown in Fig. (3.2) for a simple binary heterostructure. Although the theoretical treatment is the same for both types, we will only consider here type II heterostructures, like in the Ga\textsubscript{0.47}In\textsubscript{0.53}As and Al\textsubscript{0.48}In\textsubscript{0.52}As material system, lattice matched on InP. In type II heterostructures as the barrier and well materials are exchanged for electrons and holes, the dynamics in the conduction and in the valence bands are more decoupled. The electrons are mainly localized in the wells of the conduction band and the holes in the wells of the valence bands.

### 3.1 Multi-band model

At a high-symmetry point of the Brillouin zone represented by the reciprocal-space vector \(\mathbf{K}_0\), the wavefunction of an electron \(\Psi(\mathbf{r})\) can be expanded in the periodic parts \(u_{m\mathbf{K}_0}(\mathbf{r})\) of the Bloch function associated with the band-edges \(\varepsilon_{m\mathbf{K}_0}\):

\[
\Psi(\mathbf{r}) = \sum_m f_m^{(A)}(\mathbf{r}) u_{m\mathbf{K}_0}^{(A)}(\mathbf{r}) \quad \text{in } A, \\
\Psi(\mathbf{r}) = \sum_m f_m^{(B)}(\mathbf{r}) u_{m\mathbf{K}_0}^{(B)}(\mathbf{r}) \quad \text{in } B. \tag{3.1}
\]

We assume that the periodic parts of the Bloch waves are identical in material A and material B. This is a bad assumption if we consider different high-symmetry points for the A and B materials. However it is not a too strong assumption on the symmetry of the atomic orbitals if we consider for example the same bands of two III-V alloys. The assumption is stronger on the radial part of the functions \(u_{m\mathbf{K}_0}(\mathbf{r})\). However as we are looking to outer electrons that contributes to the chemical bond, the behavior of the radial part is very similar even for different atoms (M and P in Fig. (3.1)) since the angular momentum quantum number is the same. We therefore assume: \(u_{m\mathbf{K}_0}^{(A)} = u_{m\mathbf{K}_0}^{(B)}\). The wavefunction reads:

\[
\Psi(\mathbf{r}) = \sum_m f_m^{(A,B)}(\mathbf{r}) u_{m\mathbf{K}_0}(\mathbf{r}) \tag{3.2}
\]

where \(f_m^{(A,B)}\) have to be determined for a particular heterostructure. This expression has implicit assumptions: we assume that the heterostructure is accurately described by using a finite set of bands; in particular for III-V alloys, the Kane model can be applied. If we consider non-vanishing crystal momentums \(\mathbf{k}_A\) and \(\mathbf{k}_B\), we assume that the band-structure is well-described by the energy dispersions \(\varepsilon_{m\mathbf{K}_0}^{(A)}(\mathbf{k}_A)\) and \(\varepsilon_{m\mathbf{K}_0}^{(B)}(\mathbf{k}_B)\): the corresponding kinetic energies do not imply that we are running out of the approximation used for the band-structure (\(\mathbf{k} \cdot \mathbf{p}\), Kane model, etc.).
3.1. Multi-band model

We have assumed that the translational invariance (Bravais lattice) in the plane, where lies the direct vector \( r_\perp = (x, y) \), is not broken as the lattice constant \( a \) is the same for material A and B. The function \( f^{(A,B)}(r_\perp, z) \) can be factorized into an in-plane Bloch function \( e^{i k_\perp \cdot r_\perp} \) and a heterostructure function \( \chi^{(A,B)}(z) \):

\[
f^{(A,B)}(r_\perp, z) = \frac{1}{\sqrt{S}} e^{i k_\perp \cdot r_\perp} \chi^{(A,B)}(z)
\]

where \( S \) comes from the normalization of the Bloch waves. The in-plane Bloch function is the same in both materials. The heterostructure is therefore encoded by the \( \chi^{(A,B)} \) function. The associated in-plane \( k_\perp \) vector can formally span the whole Brillouin zone corresponding to the in-plane lattice, although it is usually limited to 1/10 of the zone [8].

In order to obtain the equations of motion for the \( \chi \) part of the wavefunction \( \Psi \), we now assume that the approximation is made in the vicinity of the \( \Gamma \)-point. (The results are applicable to other high-symmetry points) We therefore use the notations \( \varepsilon^{(A)}_{m0} \) and \( \varepsilon^{(B)}_{m0} \) for the band-edges.

The envelop function approximation terminology is now clear: the wavefunction is a product of a slow varying part \( f^{(A,B)} \) (with respect to the lattice constant of the crystal) representing the heterostructure (the typical variation-length is several monolayers) and a rapidly varying part given by the periodic part \( u_{m0} \) that varies at the scale of the atoms. The function \( f^{(A,B)}(r_\perp, z) \) is therefore called the envelop function.

We assumed that the periodic parts are the same in A and B materials: \( u^{(A)}_{mK_0} = u^{(B)}_{mK_0} \).

This implies that the interband matrix element of the atomic momentum is the same in both materials: \( P^{(A)} = -\frac{i}{m_0} \langle S|p_x|X \rangle = P^{(B)} \). As a consequence the \( \mathbf{k} \cdot \mathbf{p} \) matrix elements are the same and the Kane energy \( E_p \) is equal in both materials.

We require the continuity of the wavefunction \( \Psi(r) \) at each point \( r = (x, y, z) \). If we assume that the interface is the plane \( z = z_0 \), the linear independence of the atomic functions \( u^{(A,B)}_{mK_0} \) for material A and material B independently, forces the continuity of the envelop function at the interface point:

\[
f^{(A)}(r_\perp, z = z_0) = f^{(B)}(r_\perp, z = z_0) \quad \Rightarrow \quad \chi^{(A)}(z = z_0) = \chi^{(B)}(z = z_0),
\]

as the in-plane Bloch waves are identical for both materials.

The hamiltonian for the heterostructure illustrated in Fig.(3.1) can be written:

\[
H = \frac{p^2}{2m_0} + V_A(r) Y_A + V_B(r) Y_B
\]

where we have used the notations in [8]. The potential \( V_{A,B}(r) \) is the effective (one-electron) crystalline potential in material A or B. The functions \( Y_X \) is a step function which is unity inside the material X and zero outside.
If we let $H$ acts on the periodic functions $u_{m0}(r)$, we have:

$$Hu_{m0}(r) = \left( \varepsilon^{(A)}_{m0} Y_A + \varepsilon^{(B)}_{m0} Y_B \right) u_{m0}(r). \quad (3.6)$$

This equality is valid since the eigen-problem $Hu = \varepsilon u$ for material A (or B) is defined into a unit cell of A (or B). As we assume a perfect interface (we neglect defects, roughness, etc.) for the computation of the energy-bands, the A material ends exactly at the interface $z = z_0$, corresponding to a point of the underlying Bravais lattice. For $z \leq z_0$ we have: $Hu_{m0}(r) = \varepsilon^{(A)}_{m0} u_{m0}(r)$. The same argument can be applied for material B, and therefore Eq.(3.6) holds. We have to underline that the Bloch theorem for an infinite region was used for solving $Hu = \varepsilon u$ in each region separated by the interface located at $z = z_0$. Here we do not consider boundary effects occurring in a finite or semi-infinite regions. In the envelop function model we therefore assume infinite crystal bands even if the region is finite.

We are seeking for a set of equations that determines the $\chi_m$ functions for a finite set of bands. In a layer of a given material, the problem should read:

$$\mathcal{D} \cdot \mathbf{f} = \varepsilon \mathbf{f}, \quad (3.7)$$

where $\mathcal{D}$ is a differential operator that acts on the vector of solution $f(r_\perp, z)$ with an associated energy $\varepsilon$. For a finite set of $N$ bands we can determine the matrix elements $\mathcal{D}_{mn}$. The vector $f$ is simply equal to $(f_1, f_2, \ldots, f_N)$ where $1, 2, \ldots, N$ is the band index.

The dynamics is formulated for the envelop function $f(r_\perp, z)$ rather than for the $z$-dependent part $\chi_m(z)$ in order to avoid a ill-defined kinetic term.

The matching conditions at interfaces between materials should be addressed specifically. The boundary conditions at the limit of the region enclosing the whole heterostructure (the contacts) depend on the specific problem to be solved. Under these conditions, we can solve the band-structure of the heterostructure. These boundary conditions are discussed in (4.3) for a single period of a quantum cascade structures.

In order to determine the operator $\mathcal{D}$, we now let $H$ acts on the wavefunction $\Psi(r)$. We then project the resulting equation by a left-multiplication with $u_{n0}^*(r)e^{-i\mathbf{k}_\perp \cdot \mathbf{r}_\perp}\chi^{(A,B)}_{m0}$ followed by an integration over the space. We do not show the details of the calculation here. We refer the reader to [8]. The operator $\mathcal{D}$ reads:

$$\mathcal{D}_{mn} = \left[ \frac{\varepsilon^{(A)}_{m0} Y_A + \varepsilon^{(B)}_{m0} Y_B}{2m_0} + \frac{\hbar^2 k^2}{2m_0} - \frac{\hbar^2}{2m_0} \frac{\partial^2}{\partial z^2} \right] \delta_{mn} + \frac{\hbar k}{m_0} \cdot \langle m|p_\perp|n \rangle - \frac{i\hbar}{m_0} \langle m|p_z|n \rangle \frac{\partial}{\partial z} \quad (3.8)$$
where the evaluation of the braket $\langle m|p|n \rangle$ is obtained by integration over the unit cell $\Omega_0$ of the ordered product of the operator $p$ with periodic functions $u_{m0}$:

$$\langle m|p|n \rangle = \int_{\Omega_0} d^3r u_{m0}^*(r) p u_{n0}(r). \quad (3.9)$$

The matrix in Eq.(3.8) is the usual $k \cdot p$ matrix for a bulk material in Eq.(2.11). Excepted that the crystal momentum $k_z$ has been replaced by a differential operator:

$$k_z \rightarrow i \frac{\partial}{\partial z} \quad (3.10)$$

and that the band energies depend on the position in the crystal.

An important definition can be introduced. Instead of considering the term $\varepsilon_{m0}^{(A)} Y_A + \varepsilon_{m0}^{(B)} Y_B$, we introduce a step function $V_m$ for band $m$ defined by:

$$V_m(z) = \begin{cases} 0 & \text{for } z \text{ in material A}, \\ \varepsilon_{m0}^{(B)} - \varepsilon_{m0}^{(A)} & \text{for } z \text{ in material B}. \end{cases} \quad (3.11)$$

The function $V_m(z)$ therefore represents the band discontinuity between materials A and B. We have:

$$\varepsilon_{m0}^{(A)} Y_A + \varepsilon_{m0}^{(B)} Y_B = \varepsilon_{m0}^{(A)} + V_m(z) \quad (3.12)$$

The $V_m$ discontinuities are very difficult to determine theoretically as this requires a modeling of the whole band-structure at the interface. The $V_m$ are therefore considered as being adjustable parameters. The latter are determined by optical measurements in model systems [8].

We already know that the $\chi_m(z)$ functions must be continuous, in particular at the interface point. If we integrate Eq.(3.7) across the interface, we obtain supplementary conditions that involves the first derivative of $\chi$:

$$M^{(A)} \cdot \chi^{(A)}(z = z_0) = M^{(B)} \cdot \chi^{(B)}(z = z_0), \quad (3.13)$$

where $M^{(A,B)}$ are two $N \times N$ matrices:

$$M_{mn} = -\frac{\hbar}{2m_0} \left[ \frac{\partial}{\partial z} + \frac{2i}{\hbar} \langle m|p_z|n \rangle \right]. \quad (3.14)$$

These conditions are in general more complicated that the usual continuity of the first derivative of the wavefunction and are not very intuitive. However a simpler condition is obtained for the effective two-band model (2.7).

An additional slow-varying function $\varphi(r)$ can be added to the operator $\mathcal{D}$:

$$\mathcal{D} \rightarrow \mathcal{D} + \mathbb{I}\varphi \quad (3.15)$$
where $I$ is identity operator. This additional function can model a general electro-static potential due to an electric-field or to a self-consistent charge distribution for example.

The $\mathbf{k} \cdot \mathbf{p}$ part of $D$ can be diagonalized exactly in the Kane model, but we will only consider here the effective two-band model introduced previously. However the treatment of the general case has given us the necessary tools to derive the equations of motion of the envelop function for basically any multi-band models. Once the general differential matrix-equation in a particular band-model is obtained, the following substitutions have to be performed:

$$k \rightarrow i\nabla \quad \text{and} \quad \varepsilon_{m0} \rightarrow E_m(r),$$

where the growth direction is arbitrary.

In these substitutions the band-edges become dependent on the position $r = (x, y, z)$, as the material composition varies in space. At the end we should use the $V_m(r)$ functions introduced before. They are relevant as the important parameters are the discontinuities at the interface between materials and not the absolute energy value of a particular band. However, in order to obtain the differential equations for the envelop function, the position-dependent band-edges are more practical.

### 3.2 Envelop function formalism in the effective two-band model

For an arbitrary direction parallel to the quantization axis of the total angular momentum, the hamiltonian of the effective two-band problem introduced in (2.7) reads:

$$H_{C,V} = \left( \begin{array}{cc} E_c & \frac{\hbar m_p}{m_0} \mathbf{p}_{cv} \cdot \mathbf{k} \\ -\frac{\hbar}{m_0} \mathbf{p}_{cv} \cdot \mathbf{k} & E_v \end{array} \right).$$

(3.17)

The model involves the conduction band $E_c$ and an effective valence band $E_v$. Let’s solve the eigen-problem $H_{C,V} \psi = E \psi$, where $\psi = (\phi_c, \phi_v)$ is the solution. We have two linear equations where we perform the substitutions in Eq.(3.16) in order to obtain the equations of motion for the envelop function. This yields:

$$E_c(z) \phi_v + i \frac{\hbar}{m_0} p_{cv} \nabla \phi_v = E \phi_c, \quad \frac{\hbar}{m_0} p_{cv} \nabla \phi_c + E_v(z) \phi_v = E \phi_v.$$  

(3.18)
Both equations can be re-expressed in terms of $\phi_c$ and $\phi_v$ only:

$$\left( \frac{\hbar^2}{2m_0} \nabla \frac{E_p}{E - E_v(r)} \nabla + E_c(z) \right) \phi_c = E \phi_c$$

$$\left( \frac{\hbar^2}{2m_0} \nabla \frac{E_p}{E - E_c(r)} \nabla + E_v(z) \right) \phi_v = E \phi_v$$

(3.19)

All the assumptions on the envelop function hold. In particular $|p_{cv}|$ and therefore the Kane energy are constant in all materials. The equations for $\phi_c$ and $\phi_v$ are symmetric under $c \leftrightarrow v$ are expected. The conduction band mass $m_c$ and the effective valence mass $m_v$ are given by:

$$m_c(E, r) = \frac{1}{m_0} \frac{E_p}{E - E_v(r)}$$ and $$m_v(E, r) = \frac{1}{m_0} \frac{E_p}{E - E_c(r)}.$$ (3.20)

The mass depends on the position $r$ through the dependence of the band-edges on the position. Without non-parabolicity the mass $m(E = 0)$ depends on the considered material.

We consider heterostructures in the $z$-direction. The band-edges depend only on $z$ and therefore the mass: $m_{c,v}(E, z)$. We drop the effective valence band component as if $\phi_c$ is known then $\phi_v$ can be deduced algebraically. We can express the conduction-band mass in terms of the excess energy in the conduction band:

$$m(z, E) = m(z) \left( 1 + \frac{E'}{E_g} \right)$$ (3.21)

As shown in Eq.(2.51) this relation is exact.

The equation that determines the envelop function in the conduction band is given in Eq.(3.19). In the envelop function approximation we assume that:

$$\phi_c(r_\perp, z) = \frac{1}{\sqrt{S}} e^{i k_\perp \cdot r_\perp} \chi_c(z)$$ (3.22)

A similar expression holds for $\phi_v$. We do not write a general combination of periodic parts $u_{n0}$, as the hamiltonian, derived from the Kane model, has already been diagonalized for a subset of bands; here for C and V bands. We insert the above expression in Eq.(3.19).

This yields:

$$\left( \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_c(z, E)} \frac{\partial}{\partial z} + \frac{\hbar^2 k_\perp^2}{2m_c(z, E)} + E_c(z) \right) \phi_c(r_\perp, z) = E \phi_c(r_\perp, z)$$ (3.23)

The in-plane kinetic term $\hbar^2 k_\perp^2 / 2m_c(z, E)$ arises from the in-plane Bloch waves. For $k_\perp = 0$, the in-plane kinetic term vanishes and $\chi_c(z)$ solves an effective Schrödinger equation:

$$\left( \frac{\hbar^2}{2} \frac{d}{dz} \frac{1}{m_c(z, E)} \frac{d}{dz} + E_c(z) \right) \chi_c(z) = E \chi_c(z)$$ (3.24)
However for \( k_{\perp} \neq 0 \) the situation is more complicated as the in-plane kinetic term depends on \( z \) and \( E \). The function \( \chi_c(z) \) depends on \( k_{\perp} \) and in principle we should solve the equation for each value of \( k_{\perp} \). Here we treat this term as a perturbation. Up to the first order the solution \( \chi_c^{(0)}(z) \) obtained for \( k_{\perp} = 0 \) is the same for all values of \( k_{\perp} \). However the energy \( E \) is modified by the average value of the in-plane kinetic term:

\[
E(k_{\perp}) = E_0 + \left< \chi_c^{(0)} \left| \frac{\hbar^2 k_{\perp}^2}{2m_c(z,E_0)} \chi_c^{(0)} \right. \right> = E_0 + \frac{\hbar^2 k_{\perp}^2}{2m_c}
\]

(3.25)

where \( E_0 \) is the energy corresponding to the solution \( \chi_c^{(0)} \). The averaged mass \( \overline{m_c} \) is defined by:

\[
\frac{1}{\overline{m_c}} = \int dz \frac{1}{m_c(z,E_0)} |\chi_c^{(0)}(z)|^2
\]

(3.26)

The solution \( \phi_{c,mk_{\perp}}(r_{\perp},z) \) is a conduction subband and its energy dispersion-relation is given by \( E_m(k_{\perp}) \).

Here come some remarks about the general multi-band model developed in the previous section and the effective two-band model presented here. The effective two-band model is a simplification of the Kane model. We obtain the effective masses by a direct diagonalisation of the \( k \cdot p \) term. In the effective model the diagonalisation is achieved by solving the algebraic equations that link the conduction and the effective valence components. We obtain directly the mass \( m_c(z,E) \) for the kinetic term in the \( z \)-direction as well as for the in-plane Bloch waves. We do not consider remote bands. In the multi-band case, the \( k \cdot p \) approximation can be performed directly, or more accurately the Kane model may be considered for a band subset and the \( k \cdot p \) approximation can be possibly performed afterwards with remote bands.

In the effective two-band model the conditions on \( \chi_c(z) \) at the interface \( z = z_0 \) are simpler than in the general case. We already know that \( \chi(z) \) is continuous. If we integrate the Eq.(3.24) across the interface on \([z_0 - \epsilon, z_0 + \epsilon]\), we have:

\[
\int_{z_0-\epsilon}^{z_0+\epsilon} \left[ -\frac{\hbar^2}{2} \frac{d}{dz} m_c(z,E_0) \frac{d}{dz} + E_c(z) \right] \chi_c^{(0)}(z) = E_0 \int_{z_0-\epsilon}^{z_0+\epsilon} dz \chi_c^{(0)}(z).
\]

(3.27)

We integrate the first term on the left hand-side and we re-arrange the equation:

\[
\frac{1}{m_c(z,E_0)} \frac{d}{dz} \chi_c^{(0)} \bigg|_{z_0-\epsilon}^{z_0+\epsilon} = \frac{2}{\hbar^2} \int_{z_0-\epsilon}^{z_0+\epsilon} dz (E_c(z) - E_0) \chi_c^{(0)}(z).
\]

(3.28)

The mass is different for each material and therefore: \( m_c(z^+,E) \neq m_c(z^-,E) \). The left-hand side of the equation is apriori not zero. However if we assume that \( E_c(z) \), \( E \) and
\(\chi_c^{(0)}(z)\) are bounded on \([z_0 - \epsilon, z_0 + \epsilon]\), the right-hand side will vanish when \(\epsilon \to 0\). This yields the following continuity condition:

\[
\frac{1}{m_c(z, E_0)} \frac{d\chi_c^{(0)}}{dz} \bigg|_{z_0^+} = \frac{1}{m_c(z, E_0)} \frac{d\chi_c^{(0)}}{dz} \bigg|_{z_0^-}.
\]

We can define the first continuous derivative of \(\chi_c^{(0)}\) by:

\[
D\chi_c^{(0)} = \frac{m_0}{m_c(z, E_0)} \frac{d\chi_c^{(0)}}{dz}.
\]

Then \(D\chi_c^{(0)}\) and \(\chi_c^{(0)}\) are continuous in the whole heterostructure.

The problem of the computation of subbands in a planar heterostructure can be mapped on an effective one-dimensional problem by looking at Eq.(3.23) only. The conduction-band edge \(E_c(z)\) playing the role of a one dimensional confinement potential. This simplification is tempting but the effective valence component has an important influence on:

(a) The normalization and the orthogonality of the envelop function.

(b) The computation of oscillator strengths (and dipoles) between the conduction subbands.

These issues are addressed in the next paragraph.

### 3.2.1 Calculation of dipole moments between conduction-subbands

We assume that the envelop function is approximated with the function \(\chi_c^{(0)}\) (for \(k_\perp = 0\)), the in-plane kinetic term being treated as a perturbation. In order to avoid complicated notations, we write \(\phi_c\) the conduction-subbands.

The accurate computation of the oscillator strengths is important as the coupling between subbands and radiation is given by the dipole moments (8.1.1).

The energy-dependent mass in the effective Schrödinger equation in Eq.(3.23) has important consequences: the conduction-band states \(\phi_c\) are not orthogonal and the dipole matrix between conduction-subbands only is ill-defined as expressions like \([H, z] = ip_z/m\) have no meaning.

A correct formulation needs to be expressed in terms of the complete wavefunction: \(\psi = \)
(φc, φv). The oscillator strength \(f_{ij}\) between two subbands \(ψ^{(i)}\) and \(ψ^{(j)}\) reads:

\[
f_{ij} = \frac{2}{m_0} \left| \frac{\langle ψ^{(i)} | P | ψ^{(j)} \rangle}{E_j - E_i} \right|^2 \]

where,

\[
\langle ψ^{(i)} | P | ψ^{(j)} \rangle = -\frac{1}{2} \left\langle φ^{(i)}_c \left| \frac{m_0}{m(E_i, z)} - \frac{m_0}{m(E_j, z)} p_z \right| φ^{(j)}_c \right\rangle .
\]

(3.31)

The operator \(P\) is the momentum operator in the two band model:

\[
P = \begin{pmatrix} 0 & p_{cv} \\ -p_{cv} & 0 \end{pmatrix}
\]

(3.32)

It is worth to mention that the second effective valence state \(ψ'\) (2.7) (before the reduction to a two-band model) does not contribute to optical transitions as the matrix elements of \(P\) are identically zero for this state. The computation of the oscillator strengths in the effective two-band model are therefore exact with respect to the original three-band model (2.6).

As shown in Eq.(3.31), the braket between states \(ψ^{(i)}\) and \(ψ^{(j)}\) can be expressed by using the conduction-band component only. The conduction-band states have a special normalization that arises from the normalization of the complete wavefunction: \(⟨φ_c|φ_c⟩ + ⟨φ_v|φ_v⟩ = 1\). This condition can be translated as:

\[
⟨φ^{(i)}_c|1 + T^{(i)}|φ^{(i)}_c⟩ = 1 \quad \text{with} \quad T^{(i)} = \frac{E^{(i)} - E_c(z)}{E^{(i)} - E_v(z)} ≡ \frac{ε^{(i)}(z)}{ε^{(i)}(z) + E_g(z)},
\]

(3.33)

where in the last equivalence the definition \(E^{(i)} = ε^{(i)}(z) + E_c(z)\) was used, while \(E_g(z) = E_c(z) - E_v(z)\) is the effective gap in the two-band model. This reformulation will be useful in the next chapter and we already introduce it here.

The oscillator strength \(f_{ij}\) is related to the dipole moment \(⟨ψ^{(i)}|Z|ψ^{(j)}⟩\), where \(Z\) is the position operator defined by the commutation relation \(P = im_0\hbar^{-1} [H, Z]\). However the dipole moment \(Z\) is ill-defined for infinite systems. The oscillator strength can be written:

\[
f_{ij} = \frac{2m_0}{\hbar^2} (E_j - E_i) \left| \langle ψ^{(i)} | Z | ψ^{(j)} \rangle \right|^2.
\]

(3.34)

The absolute value of the dipole moment \(|Z_{ij}| = \left| \langle ψ^{(i)} | Z | ψ^{(j)} \rangle \right|\) can therefore be evaluated in terms of the conduction-band states \(φ_c\):

\[
|Z_{ij}| = \frac{\hbar}{2m_0} \frac{1}{(E_j - E_i)} \left| \left\langle φ^{(i)}_c \left| p_z \frac{m_0}{m(E_i, z)} \right| φ^{(j)}_c \right\rangle - \left\langle φ^{(i)}_c \left| \frac{m_0}{m(E_j, z)} p_z \right| φ^{(j)}_c \right\rangle \right|.
\]

(3.35)
In the position representation where the wavefunction is $\phi_c$ and its first continuous derivative is $D\phi_c$; we have:

$$|Z_{ij}| = \frac{\hbar^2}{2m_0 |E_j - E_i|} \left| \int dz \left( D\phi_c^{(i)} \phi_c^{(j)} - \phi_c^{(i)} D\phi_c^{(j)} \right) \right|. \quad (3.36)$$

This expression will be very useful for numerical evaluations. It is important to remark that the states $\phi_c^{(i,j)}$ need to be normalized using the special normalization in Eq.(3.33). On the other hand, the oscillator strength can be computed from numerical values of the dipole moment by using Eq.(3.34). The scaled oscillator strength $f' = (m^*/m_0)f$ cannot be defined here since we are considering materials with different effective masses.
Chapter 4

The band-structure of quantum cascade lasers

The numerical implementation of the effective two-band model is presented in (4.2) for bound-states as well as for continuum states. The computation of the band-structure of a typical quantum cascade laser is performed in (4.3). There we discuss the structure with and without electric field, showing that the same band-solving method can be used for both. We then sketch conceptually how a transport model should be implemented (4.3.3) and what quantities should be necessarily computed from the band-structure. In (4.4) we develop a tight-binding model. The latter will allow the computation of the coupling energies between states coupled by barrier potential regions. These are key quantities in the modeling of coherent effects. The precise way of evaluating practically the coupling energies is shown (4.4.3). The model is then briefly applied to a quantum cascade structure (4.4.4). Finally we develop a model for computing the self-consistent potential in the Hartree approximation (4.5) for a periodic structure. The case of a thermal distribution is discussed (4.5.5). The general algorithm for solving the self-consistent potential in the various cases considered in this work is then presented (4.5.6).

4.1 Formulation of the problem for energy-eigenstates

As discussed briefly in the introduction, the energy-eigenstates of a planar heterostructure under an applied electric field are not exactly bound-states but resonances with a natural width given by their coupling with the continuum of states in the ionization region of the spectrum. In Fig.(4.2) the bound-state region and the continuum region are illustrated for a three-quantum-well system. Commonly the quantum cascade structures are designed
so that most of the energy-eigenstates are deeply bound into the structure, their energies being below the band discontinuity. However the upper laser-state can be pushed near the ionization energy-region in the conduction-band for short wavelength mid-infrared structures. [29,30,87,118,120] The coupling to the continuum cannot be neglected. Nevertheless in all the models we will present here, the resonances are approximated with bound-states. The impact of the continuum will be taken into account by computing the optical absorption between bound-states and the continuum and by allowing an escape time $\tau_{\text{esc}}$ in rate equation models. The latter may be important as the coupling of the upper laser state to the continuum reduces the population inversion.

In the introduction we have already outlined that the modeling of quantum effects, like resonant tunneling, require localized states. In quantum models [61,62,64,109,119] this requirement is not necessary because of the representation invariant formulation of the problem. In effective models it becomes important since perturbative approximations require a particular representation choice, like two localized states coupled by a barrier.

For solving the band structure, we use the effective two-band model discussed in (3.2). It gives eigen-energies that corresponds to subband-edges in the conduction band. This model is described in [89]. When the problem is reduced to a one-band one-dimensional system with an effective mass $m^*(z)$, the model only involves orbitals with an $s$-symmetry. This approximation holds for energies that are close to the bottom of the band. When the energies become comparable to the gap energy, the valence band has to be taken into account with the Kane model possibly reduced to an effective two-band model.

As we aim to develop a predictive transport model we need to use a band model that gives numbers in a tight agreement with experimental measurements, as the effective two-band model. The latter incorporates the minimal relevant informations that gives accurate predictions. Its internal parameters as the non-parabolicity coefficient $\gamma$, the $k \to 0$ masses and the band-edge discontinuities $V_c$ are parameters that are determined and affined with measurements.

We can give a more practical expression for the energy-dependent mass, already used in Eq.(3.21). We rewrite $E = \epsilon(z) + E_c(z)$, $\epsilon$ being the excess energy in the conduction band. Then $m^*(z, E)$ reads:

$$m(z, \epsilon) = m_0 \frac{\epsilon + E_v(z) - E_v(z)}{E_p} \quad (4.1)$$

For a binary heterostructure we can obtain a practical expression for the mass. We put in evidence $E_g = E_c(z) - E_v(z)$. The Kane energy can be expressed by using the well parameters as $E_p = E_{g,w}/m^*_w$ where the subscript $w$ holds for the well in the expression $E_{g,w}$. In the same way $E_{g,b}$ will be the effective gap for the barriers. In the wells the
4.1. Formulation of the problem for energy-eigenstates

the energy-dependent mass reads:

\[ m(z, \epsilon) = m_0 m_w^* \left( 1 + \frac{\epsilon}{E_{g,w}} \right) \]  
for \( z \) in the wells.  \( (4.2) \)

A similar expression holds inside the barriers:

\[ m(z, \epsilon) = m_0 m_w^* \left( \frac{E_{g,b}}{E_{g,w}} + \frac{\epsilon}{E_{g,w}} \right) \]  
for \( z \) in the barriers.  \( (4.3) \)

By assuming the same Kane energy (that was a basic requirement for the envelop function approximation) in the well and in the barrier, we have:

\[ \frac{E_{g,b}}{E_{g,w}} = \frac{m_b^*}{m_w^*}. \]  \( (4.4) \)

In the barriers:

\[ m(z, \epsilon) = m_0 m_b^* \left( 1 + \frac{\epsilon}{E_{g,w} m_b^* / m_w^*} \right). \]  \( (4.5) \)

At the end we rewrite the energy-dependent effective mass with a gap function \( E_G(z) \):

\[ m(z, \epsilon) = m^*(z) \left( 1 + \frac{\epsilon(z)}{E_G(z)} \right) \]  
with \( E_G(z) = E_{g,w} \frac{m^*(z)}{m_w^*}. \)  \( (4.6) \)

This last expression is very useful in the numerical evaluation of the mass. The position-dependent effective mass \( m^*(z) \) is the effective mass evaluated at the conduction-band edge.\(^1\)

The effective mass depends linearly on the energy \( \epsilon \). In wells and in barriers the slope is the same \( m_w / E_{g,w} = m_b / E_{g,b} \). In lattice matched Ga\(_{0.47}\)In\(_{0.53}\)As /Al\(_{0.48}\)In\(_{0.52}\)As on InP, this slope in units of the bare electron mass \( m_0 \) is 0.054 eV\(^{-1}\). The effective mass in the wells is typically \( m_w = 0.0427 m_0 \) and in the barriers \( m_b = 0.076 m_0 \). The non-parabolicity coefficient is \( \gamma = 1.13 \cdot 10^{-18} \text{ m}^2 \). As the conduction band offset between wells and barriers is \( \Delta E_c = 0.52 \text{ eV} \), the variation of the effective mass with respect to the energy cannot be neglected. When one probe an electron in a well region at an energy of 0.4 eV, it is 1.5 times heavier than an electron in the bottom of the well. This has important consequences on the allowed energies and the scattering rates, as subbands largely separated in energy will have energy dispersions with very different curvatures.

The energy-dependent effective mass Eq.(4.6) is valid only when the kinetic energy is largely smaller than the effective gap. The energy \( \epsilon = E - E_c(z) \) can be negative in

\(^{1}\)The above formula has been directly implemented in sewlab, so the effective gap in the input file of the structure is the function \( E_G(z) \) evaluated in the wells and in the barriers for a binary heterostructure. The latter being typically evaluated by the sewself code in the beginning.
barriers. In the latter case, when the energy approach the energy of the gap $E_G$ (0.79 eV in wells and 1.41 eV in barriers for Ga$_{0.47}$In$_{0.53}$As /Al$_{0.48}$In$_{0.52}$As), the effective mass becomes zero or negative, which has no physical meaning here. The range of validity in energy for $m^*(\epsilon)$ can be extended by considering quadratic or higher-level powers of $\epsilon$. However a more efficient way of solving this problem, as least partially, is to apply directly the Kane model (without further reductions) in the envelop function approximation.

4.2 Numerical implementation of the solution

4.2.1 General explanation of the method

In order to enable the treatment of an electric field or a self-consistent potential, we add an extra electro-static potential $q_0\Phi(z)$ to the effective Schrödinger equation in Eq.(3.24):

$$
\left(p_z^2 m(z,E)p_z + V(z) + q_0\Phi(z)\right) = E\phi_c \quad \text{with} \quad m(z,E) = m^*(z) \left(1 + \frac{\epsilon(z)}{E_G(z)}\right)
$$

where $\epsilon(z) = E - V(z)$ and $E_G(z)$ is the gap function we have expressed previously for a binary heterostructure. The potential $V(z)$ has formally replaced the conduction band-profile $E_c(z)$ in order to give a formulation closer to the usual problem of an electron in a one-dimensional potential.

As the case of infinite systems is important for the foundation of a complete quantum model, we will only consider here finite potential regions. Moreover we will implicitly restrict ourselves to a binary heterostructure by choosing the special gap function given above. Before we give numerical results for quantum cascade structures, we examine the numerical problem by keeping an arbitrary energy-dependent mass $m(z,E)$.

As the mass depends on the energy, the usual finite-element methods [71] used for solving a typical eigen-problem $H\phi = E\phi$ will not work, since the operator $H$ is hermitian and does not depend on energy. It is necessary to use a shooting method [77] with a convergence criterion on the eigen-function in some extra convergence-region.

Let us assume a finite region $D = [0, L]$, where $L$ is the length of the potential region. In $D$ the potential is given by $V(z)$ and the kinetic term involves the energy and position dependent mass $m(z,E)$. We discretize the space $D$ in $N$ subsets $D_i$ so that $\bigcup_{n=1}^{N} D_i = D$. The subsets $D_i$ will only share single points (interfaces) between adjacent domains: $D_i \bigcap D_j = \partial D_i \bigcap \partial D_j \neq 0$ iff $|i - j| = 1$.

As we are modeling heterostructures that consist in by a succession of layers of different materials and of different thicknesses, we can choose an efficient mesh for $D$. We assume
4.2. Numerical implementation of the solution

that the potential region is a succession of $M$ layers, of thicknesses $T_i$ with $i = 1..M$. The layers are arranged so that the first layer occupies the space $]0, T_1]$ and the second layer the space $]T_1, T_1 + T_2]$, until the domain $D$ is filled. The interface points are not included and will be treated separately. We write the boundary of the layers with the notation $X_k = \sum_{j=1}^{k} T_j$. The layers are arranged in domains $\Lambda_i \equiv ]X_{i-1}, X_i[$ with $i = 1..M$. The boundary of the region being given by $X_0 = 0$ and $X_M = \sum_{j=1}^{M} T_j \equiv L$. Therefore $[X_0, X_M] \equiv D$.

In each domain $\Lambda_i$ the material is the same with a constant potential $V_i$ (we do not have incorporated the electro-static potential $e\Phi(z)$ in the potential function $V(z)$) and an energy-dependent mass $m_i(E)$.

In the case of a vanishing electro-static potential ($q_0\Phi = 0$), the differential equation Eq. (4.7) can be solved analytically in each domain $\Lambda_i$. Since $[p_z, 1/m_i(E)] = 0$ for $z \in \Lambda_i$, the eigen-equation reads:

\[
\left( \frac{d^2}{dz^2} + 2m_i(E)(E - V_i) \right) \phi_c(z) = 0. \tag{4.8}
\]

This is a second-order differential equation which admits two types of solutions: the first is oscillatory when $E > V_i$, the second is exponential when $E < V_i$. There is also a limit case when $E = V_i$. This special case should be considered as the limit of the two other cases in order to avoid an ill-defined solution: the wavefunction is constant. The oscillatory solution reads:

\[
\phi_i(z) = A_i \cos(k_iz) + B_i \sin(k_iz) \quad \text{with} \quad k_i = \sqrt{2m_i(E)(E - V_i)}, \quad \text{for} \quad E > V_i. \tag{4.9}
\]

The exponential solution reads:

\[
\phi_i(z) = A_i \exp(\kappa_iz) + B_i \exp(-\kappa_iz) \quad \text{with} \quad \kappa_i = \sqrt{2m_i(E)(V_i - E)}, \quad \text{for} \quad E < V_i. \tag{4.10}
\]

Where $A_i$, $B_i$ are the coefficient for each orthogonal solutions in both cases. We have dropped the conduction-band index $c$ in order to lighten the notation.

The wavefunction in each domain $\Lambda_i$ is therefore represented by a vector of coefficients $(A_i, B_i)$. At interface points between two adjacent domains the continuity of the wavefunction $\phi(z)$ and of it’s first continuous derivative $D\phi(z)$ is imposed. Between domain $\Lambda_i$ and $\Lambda_{i+1}$, we impose:

\[
\lim_{z \to X^-_i} \phi_i(z) = \lim_{z \to X^+_i} \phi_{i+1}(z) \quad \text{and} \quad \lim_{z \to X^-_i} D\phi_i(z) = \lim_{z \to X^+_i} D\phi_{i+1}(z) \tag{4.11}
\]

The ± sign means that the value of $\phi_i(z)$ or $D\phi_i(z)$ is given by a limit, where the interface $X_i$ is approach from higher (+) values or from lower (−) values.
This transfer method \cite{77} works accurately. It allows to compute the solution of the equation of motion at a given energy up to a normalization constant. A convergence criterion has to be applied at the end of the propagation (at and beyond the boundary $X_M$) in order to select the bound-states.

Here consider an alternative method for computing the propagation of the wavefunction in a layered medium. This method is less power consuming in terms of evaluation of trigonometric or exponential functions and does not require to treat explicitly all the matching cases between domains with oscillatory or exponential solutions.

First of all, we consider an arbitrary domain $\Lambda$ of thickness $T$ with material parameters $m = m(E)$ and $V$. The latter yields a propagation wavevector of $k = \sqrt{2m|E-V|}$ that holds both in oscillatory and exponential cases thanks to the absolute value $|E-V|$. The origin of the domain can always be chosen at $z = 0$. At the origin of the domain, the numerical value of the wavefunction $\phi(0)$ and of its first continuous derivative $D\phi(0)$ are related to the coefficients $A$ and $B$ by simply evaluating Eq. (4.9) and Eq. (4.10) at the origin of the domain. The coefficients read:

\begin{align}
A &= \phi(0) \quad B = \frac{1}{m_0} \frac{m}{k} D\phi(0) \quad \text{(oscillatory)} \quad (4.12) \\
A &= \frac{1}{2} \left( \phi(0) + \frac{1}{m_0} \frac{m}{k} D\phi(0) \right) \quad B = \frac{1}{2} \left( \phi(0) - \frac{1}{m_0} \frac{m}{k} D\phi(0) \right) \quad \text{(exponential)} \quad (4.13)
\end{align}

The coefficient in the domain $\Lambda$ are linked to the value of the wavefunction at the left interface of the domain.

The value of $\phi$ and $D\phi$ at the right interface of the domain can be evaluated directly. This gives the wavefunction value from the coefficients:

\begin{align}
\phi(T) &= A \cos(kT) + B \sin(kT) \quad D\phi(T) = m_0 \frac{k}{m} (-A \sin(kT) + B \cos(kT)) \\
\phi(T) &= A e^{kT} + Be^{-kT} \quad D\phi(T) = m_0 \frac{k}{m} (Ae^{kT} - Be^{-kT}) \quad (4.14)
\end{align}

It is also relevant to be able to evaluate $\phi$ and $D\phi$ at any point $z$ inside the domain $\Lambda$ in terms of the values $\phi(0)$ and $D\phi(0)$ at the left boundary of the domain\footnote{The equations for the exponential case are not numerically efficient; the cosh and sinh functions should be expanded in exponentials in order to reduce the computational effort.}:

\begin{align}
\phi(z) &= \phi(0) \cos(kz) + \frac{1}{m_0} \frac{m}{k} D\phi(0) \sin(kz) \quad D\phi(z) = -m_0 \frac{k}{m} \phi(0) \sin(kz) + D\phi(0) \cos(kz) \quad (4.15)
\end{align}
\[ \phi(z) = \phi(0) \cosh(kz) + \frac{1}{m_0} m \phi(0) \sinh(kz) \quad D\phi(z) = m_0 \frac{k}{m} \phi(0) \sinh(kz) + D\phi(0) \cosh(kz) \]

The above expressions, coefficients from wavefunctions at the left boundary, wavefunctions from coefficients at the right boundary and propagation of the wavefunctions inside the domain from the values of the wavefunctions at the left boundary of the domain, are the three necessary and sufficient operations that allow to compute the coefficients \((A, B)\) in any layered medium, together with the wavefunction and its first continuous derivative at any point in the medium. We are considering propagation from left to right. Similar expressions are available for propagation from right to left.

We now consider again the domain \(D\) tiled by material layers in sub-domains \(\Lambda_i \equiv ]X_{i-1}, X_i[\) with \(i = 1..M\). We assume initial coefficients \(A_1\) and \(B_1\) in the first layer \(\Lambda_1\). The propagation is performed with the following steps:

- (1) The wavefunctions \(\phi(X_0)\) and \(D\phi(X_0)\) are evaluated at the left boundary \(X_0\) of the first domain with initial material parameters \(m^{(0)}\) and \(k^{(0)}\) and initial coefficients \(A_1\) and \(B_1\) using Eq.(4.14) and Eq.(4.15) with \(T = 0\).

- (2) The wavefunctions are propagated to the right boundary \(X_1\) of the first domain \(\Lambda_1\) using Eq.(4.16) and Eq.(4.17) with \(z = X_1 - X_0\) (the diameter of domain \(\Lambda_1\)) and the parameters \(m^{(1)}\) and \(k^{(1)}\) of the first layer. This gives the values \(\phi(X_1)\) and \(D\phi(X_1)\) from values \(\phi(X_0)\) and \(D\phi(X_0)\).

- (3) The coefficients \(A_2\) and \(B_2\) of the adjacent domain \(\Lambda_2\) are obtained using Eq.(4.12) and Eq.(4.13) with the values of the wavefunctions at the left interface of \(\Lambda_2\) corresponding to the right interface of \(\Lambda_1\), with wavefunctions values \(\phi(X_1)\) and \(D\phi(X_1)\).

- (4) The propagation is chained to the next domain \(\Lambda_3\) by propagating the wavefunctions across domain \(\Lambda_2\), with \(z = X_2 - X_1\), giving values \(\phi(X_2)\) and \(D\phi(X_2)\), as in step (2).

- The propagation stops when the wavefunctions values are obtained at the right interface of the last domain \(\Lambda_M\) with \(\phi(X_M)\) and \(D\phi(X_M)\).

After the propagation, the values of the coefficients are known in each domain \(\Lambda\) with: \((A_1, B_1), (A_2, B_2), .. (A_M, B_M)\). The wavefunction values are also known at each boundary: \((\phi(X_0), D\phi(X_0)), (\phi(X_1), D\phi(X_1)), .. (\phi(X_M), D\phi(X_M))\). The wavefunctions values are indeed less important than the coefficient values, as the latter allow to know the exact values of \(\phi\) and \(D\phi\) at any point of the domain \(D\), including boundaries.
We now address the question of the convergence for bound-states and the adequate choice of initial coefficients. For bound-states we need to add two extra convergence-layers.

For the first we add an extra domain \( \Lambda_0 \) before the layer sequence with parameters \( m^{(0)} \) and \( k^{(0)} \). The diameter of the domain is \( \Delta_L \), thus \( \Lambda_0 = [X_0 - \Delta_L, X_0] \). We start the propagation in this layer with initial coefficients \( A_0 \) and \( B_0 \). The next coefficients \((A_1, B_1)\) are determined by the propagation algorithm. For bound-states we keep only growing exponentials in the first layer, in the direction of propagation. The initial coefficients are therefore: \( A_0 = 1 \) and \( B_0 = 0 \). The choice of \( A_0 = 1 \) is arbitrary. The solution wavefunction \( \phi \) needs to be normalized afterwards using Eq.(3.33). Indeed if the propagation energy gives an oscillatory solution in \( \Lambda_0 \) the problem is ill-defined and the propagation should not be performed.

For convergence at the end of the structure, we need to add a second convergence domain \( \Lambda_{M+1} = [X_M, X_M + \Delta_R] \) at the end of the structure. The diameter of this domain is \( \Delta_R \). The solution needs to be exponential in this domain. We want only decaying exponential in the direction of propagation for a bound-state. The propagation coefficients \( A_{M+1} \) and \( B_{M+1} \) are obtained through the values of \( \phi(X_M) \) and \( D\phi(X_M) \) at the right boundary of the domain \( \Lambda_M \), using Eq.(4.12) and Eq.(4.13) with the parameters \( m^{(M+1)} \) and \( k^{(M+1)} \).

We then need to define a merit function on the growing exponential coefficient \( A_{M+1} \) in order to select bound-states. The value of \( A_{M+1} \) is expected to cross zero each times a bound-energy is reached. However searching the zeros of \( A_{M+1} \) is less efficient than bracketing the divergences of \( 1/A_{M+1} \) with a bisection technique. The perfect case where \( A_{M+1} = 0 \) needs to be implemented.

**Numerical application: two quantum-well example**

In Fig.(4.1) we show the squared moduli of the bound-states wavefunctions vertically shifted at the corresponding eigen-energy for a simple two quantum-well system considered in [89]. The layer sequence is in Å, \( T_1 = 59 \) for the thick well, \( T_2 = 13 \) for the coupling barrier and \( T_3 = 24 \) for the thin well. One each side a thick barrier of \( \Delta_L = \Delta_R = 60 \) Å is placed as convergence layers. The material system is \( \text{Ga}_{0.47}\text{In}_{0.53}\text{As} / \text{Al}_{0.45}\text{In}_{0.52}\text{As} \) lattice matched on InP. The conduction-band offset is 0.52 eV and the minimum of the energy is \( E = 0 \) in the bottom of the wells.

The energy region \( E \in [E_{\text{min}} = 0, E_{\text{max}} = 0.52] \) eV is limited to the bound states. The energy-region is split into \( N \) equivalent sub-domains. \( N \) needs to be chosen sufficiently large in order to have only one bound-state per sub-domain. The minimal splitting between two eigen-energies needs to satisfy: \( |E_{i+1} - E_i| < |E_{\text{max}} - E_{\text{min}}|/(N - 1) \). This becomes
important with thick barriers. The propagation is performed for each energy: \( E_k = E_{\text{min}} + (k-1)|E_{\text{max}} - E_{\text{min}}|/(N-1) \), with \( k = 1..N \). Then in each sub-domain \([E_k, E_{k+1}]\), with \( k = 1..(N-1) \), a bisection algorithm allows to find the bound-energy if it exists. A criterion on the precision of the eigen-energies needs to be set. At the end the bound-spectrum is found with eigen-energies: \( E_0, E_1, .., E_P \), assuming that the dimension of the spectrum is \( P \).

### 4.2.2 Numerical solution with an electro-static potential

The solving technique for bound-states in a layered medium for vanishing electro-static potential \( q_0 \Phi = 0 \) can be extended to problems where \( q_0 \Phi \) is an arbitrary slow-varying function. The first interesting case is an applied constant and uniform electric field \( F \). The potential takes the form \( \Phi = -Fz \) which adds a linear contribution to the staircase potential due to the layer sequence. The minus sign is for electrons, as we have written \( q_0 \Phi \) with the signless elementary charge.

![Graph showing the bound eigenstates for a two quantum-well region.](image)

Figure 4.1: The bound eigenstates for a two quantum-well region are shown. (Left) The squared modulus for each wavefunctions shifted vertically by the eigen-energy (in meV) are shown in the potential region. (Right) The corresponding squared merit function is plotted in logarithmic scale. The merit function exhibits sharp divergences at each bound-energy. The material system is Ga\(_{0.47}\)In\(_{0.53}\)As / Al\(_{0.48}\)In\(_{0.52}\)As lattice matched on InP. The layer sequence from left to right in nanometers is: 6.0, 5.9, 1.3, 2.4, 6.0. Barriers are in bold, wells in roman.
Without choosing a particular function for $q_0\Phi$, we can use the same technique as for unbiased potentials, by adding supplementary sub-domains in each $\Lambda_i$ domains to take into account the variation of the potential caused by the electro-static potential inside each layers. The sub-domain sampling needs to be high enough to model correctly the variations of $q_0\Phi$. The choice of the sampling is related to the desired accuracy on the bound-energies. More sophisticated techniques use a conditional mesh that guarantee a given accuracy. These techniques require to solve the eigen-spectrum for varying mesh conditions. In this work, we will not consider these techniques, although different kind of meshes (random, constant step, subdivision) in domains have been tried without providing a significant variation of the accuracy on the energies.

It is worth here to mention a point of detail: the electro-static potential is defined in the potential region $D = \bigcup_{k=1}^{M} \Lambda_k$, excluding the convergence layers $\Lambda_0$ and $\Lambda_{M+1}$. In the latter, $q_0\Phi$ needs to be defined by a special requirement. It can be the periodicity, when we consider a self-consistent potential, or extrapolation (typically for an applied electric field $F$), or by padding: $q_0\Phi$ being set to a constant value, usually zero. In any case this choice can alter the energy spectrum. In this work we will take care of always choosing the convergence layers and the value of $q_0\Phi$, so that these boundary layers mimic the surrounding structure: for example the adjacent periods, or the rest of the period if we solve only a portion of the latter. The inspection of each possible cases has been performed when we have implemented the numerical methods. However such technicalities, although being important for getting accurate results, is beyond the scope of this work. We only give the main steps that generalize the propagation method.

For each domain $\Lambda_k$, including the convergence layers, we consider a refinement with adjacent sub-domains $\Lambda_k^m$. We have $\Lambda_k = \bigcup_m \Lambda_k^m$. We give a size limit on the diameter: diameter ($\Lambda_k^m$) $\leq W_s$, where $W_s$ is the discretization step in the layer. $W_s$ has to be chosen small enough in order to keep a good accuracy. In each $\Lambda_k^m$ the material parameters are constants with $m^{(k)}(E)$, $V^{(k)}$ and $q_0\Phi(k,m)$. The energy-dependent mass and $V^{(k)}$ value do not depend on the refinement as they are constant in each layers. In $\Lambda_k^m$ we end with similar equations as Eq.(4.8) but with a modified propagation constant $k^{(k,m)} = \sqrt{2m^{(k)}|E - V - q_0\Phi(k,m)|}$. The value of $\Phi^{(k,m)}$ can be the average of boundary values $\{\Phi(\partial \Lambda_k^m)\}$ or the left/right boundary values. At the end, we have a similar problem than with a vanishing electro-static potential but with adjacent domains $\Lambda_k^m$ tiling the potential region and the convergence layers. The propagation technique can therefore be directly applied.
4.2. Numerical implementation of the solution

When the structure is biased with an electric field, we need to add extra box-layers. The latter can be very thin, a few angströms. They surround the structure and the convergence layers in order to allow to define correctly the merit function on an energy range that covers all the possible bound energies from $E_{\text{min}} = \min_{z \in \Lambda_0 \cup D \cup \Lambda_{M+1}} (V(z) + q_0\Phi(z))$ to $E_{\text{max}} = \max_{z \in \Lambda_0 \cup D \cup \Lambda_{M+1}} (V(z) + q_0\Phi(z))$. The box-layers are barriers with at least the maximal potential height of $E_{\text{max}} - E_{\text{min}}$. They are typically made of the same material used for the barriers in the potential region. They not only allow to define correctly the merit function but can also trigger resonances (that are unbounded to the potential) in the
continuum above the quantum wells. The continuum states are approximated with bound-states over the total length of the potential region. Such approximation is indeed better when the box is larger although the sum-rule developed in [89] prevent us to over-estimate the oscillator strengths between bound-states and the continuum. Fig.(4.2) summarizes the computation of bound-states with an electric field and the effect of the box-layers on the computation of the continuum states. The examined structure uses a Bragg-reflector made of two thin quantum well in order to bound a resonance to a larger quantum well. This creates a gap in the continuum above the Bragg-reflector and this method will be used in quantum cascade lasers in order to avoid leakage from the upper laser-state to the continuum. The computation of the continuum spectrum allows us to model the contribution to the optical absorption from bound to continuum transitions.

4.3 Band-structure of a quantum cascade laser

We apply the calculation of energy-eigenstates to a quantum cascade structure. We will examine the quantum cascade structure published in [9].

4.3.1 Band-structure of the period at zero-field

In Fig.(4.3) the band-structure for one period is shown. The period is usually defined by the injection barrier. The band-structure is computed at zero field, which helps to understand the design. In Fig.(4.3b) the structure is cut at the extraction barrier (22 Å) separating the active region from the injector region. The states are clearly identified: state 4 is the upper laser state. The optical transition occurs between state 4 with state 3. The lower states 3, 2 and 1 build two resonances matched at the LO-phonon energy. For the latter we use the bulk value of 32 meV in Ga0.47In0.53As wells. This active region consists of 4 quantum wells. This design is a 4 QW design or two-phonon design. The latter is referring to the method used for the depletion of the lower laser-state.

The states in the active and the injector regions are separated by an extraction (or exit) barrier. The active and the injector region cannot be more split. Even if we may split the structure at each barrier, this will not give accurate results. We can consider that the current density computed by a transport model is a serie in terms of characteristic energies. If all the terms are taken, the solution is exact and does not depend on the representation. If we truncate the expansion to the first order, the choice of the representation is crucial in order to have an approximation closest to the exact solution. When we consider effective transport models we try to choose the representation in line with the approximations.
4.3. Band-structure of a quantum cascade laser

Figure 4.3: One period of the quantum cascade structure from [9] is shown at zero electric field. (a) The propagation of the wavefunctions was performed across the whole period. Even at this field the design of the structure appears clearly: it consists of an active region and of an injector region coupled by an injection barrier of 4.0 nm and an extraction barrier of 2.2 nm. The shaded pink area in the injector region represents the doped layers with a volume concentration of Si of $2 \times 10^{17}$ cm$^{-3}$. (b) The same structure is shown, but the wavefunctions were propagated in the active region and in the injector independently. The eigenstates in the active region are labeled from 1 to 4. The level 4 is the upper laser-state. The laser transition occurs between 4 and 3 (red arrow). The lower states yield two resonances (green arrow) designed at the LO-Phonon energy of 32 meV in Ga$_{0.47}$In$_{0.53}$As wells lattice matched on InP.
4.3.2 Alignment electric field

In Fig.(4.4) a constant and uniform electric field of 48 kV/cm is applied on the structure. This field corresponds roughly to the alignment between the ground-state of the previous period with the upper-laser state of the adjacent next period. This field is sometimes called alignment field as it is typically the field where the current density is expected to be maximal. The ground-state of the injector being by design the most populated state and the most coupled state with the upper laser-level (its amplitude of probability is maximized near the injection barrier). This field is often connected with the negative-differential-regime (NDR) since the latter is expected to take place for higher fields. This approximation is too crude in general. The coupling between the previous period and the next one will be found to not only rely on the ground-state but also on excited states of the injector region. In simulation of the current-voltage curve, the maximal current is usually reached at lower fields than the alignment field. Here, we have kept this field as a reference field as it depends only on the band structure and not on a particular transport model. Usually this is the field at which the structure is designed and optimized.

In Fig.(4.4), one period and an additional adjacent active region (on the right) are shown. The figure shows all the alignments that occur in the structure at the designed operation field. While state 4’ is aligned with the ground state G of the injector region, the states 3,2,1 are in resonance with the extraction states A,B,C of the injector region. These alignments occur at the same time in order to optimize the population inversion and therefore the optical power: the upper laser-level is efficiently populated by electrons from the injector region and the extraction from the lower laser-state and states 2,1 is maximized.

4.3.3 Conceptual transport model

Here we sketch the operation of the structure. We will examine it in details when we will discuss the transport models (6). We first consider the active region. The latter is confined between coupling barriers (injection/extraction barriers). We can examine it independently (without injection). The LO-phonon interaction tends to restore thermal equilibrium. Thus higher in energy are the states, less they are populated. Moreover the resonances between states $3 - 2$ and $2 - 1$ are matched on the LO-phonon bulk energy (32 meV). The thermal equilibrium is therefore restored as quick as possible as the momentum of the phonons is essentially zero. We neglect phonon modes confined by the heterostructure. These special scattering mechanisms will not be considered in this work. If we now couple the active and the injector regions, the injector ground state (or more
4.3. Band-structure of a quantum cascade laser

Figure 4.4: The two-phonon structure is shown at alignment field of 48 kV/cm. One period and the adjacent active period (at the right) are shown. The active region states are labeled 4,3,2 and 1. The extraction states in resonance with 3,2,1, are labeled A,B,C. The states of the right active region are labeled with a prime. The ground state of the injector region is labeled with letter G. At this field G is aligned with state 4’. The shaded orange area in the injector region covers the miniband (allowed states), while the grey area is the minigap (forbidden states): this ensures the selectivity of the injector.

precisely the injector miniband) injects electrons selectively in the upper laser-level. As the energy of the main laser transition 141 meV is largely off from the phonon energy, the scattering time from the upper laser-level is largely higher than the scattering time of the lower laser-level 3. The shape of the wavefunction is engineered around the thin barrier of 7 Å, in order to keep a large dipole of 30 Å, while the scattering time from the upper laser-state us increased. When an electron is scattered out from the upper laser-level in state 3, the thermal equilibrium is as quick as possible restored between states 3,2 and 1,
while the resonances 3-A, 2-B and 1-C extract electrons to the injector region.

In the injector region, the electrons are thermalized. In this region a common fermi-level can even be defined in many cases. The state G has the highest population: only a small fraction (10%) of the electrons are injected in the next upper laser-level.

The basic requirements of a transport model can be enumerated. We need to know the band-structure and to be able to split it into fundamental regions that cannot be split anymore without yielding a very bad approximation for scattering times. The latter being computed by the Fermi Golden rule.

We need to know the coupling energies between the fundamental sub-periods (4.4), in order to model coherent effects between them. The scattering rates inside each region need to be computed between each pair of states for each relevant interactions. The dephasing time needs to be computed across the coupling barriers in order to evaluate the current.

### 4.3.4 Injector selectivity

The injector is a central piece in the operation of quantum cascade lasers. While it stabilizes the electrical operation of the structure, it enables the selective extraction of the electrons from the active region and achieve thermalization in the ground-state G. This tailored selectivity avoid the electrons from the upper-laser state to be scattered out directly to the injector or to the continuum. The injector is designed for yielding a miniband\(^3\) in front of the state 3,2,1 and to yield a minigap in front of the upper-laser state. The design is similar to the quantum-well resonances in (4.2.3). These properties are better observed on the transparency of the injector region under alignment field.

These calculations are interesting for the design and the understanding of quantum cascade lasers. However we will not develop the detailed calculations here. The latter can be performed with the same propagation algorithm used for computing the bound-states. We have represented in Fig.(4.4) the miniband and the minigap zones by shaded zones.

\(^3\)The term *miniband* is not very accurate to describe the injector band-structure as the structure is finite and is not a superlattice. We may use the term *electron filter* but the latter is lack of completeness. Even if the injector is not a superlattice, the property of having an allowed energy range, followed by a forbidden one, is clearly visible even in a small structure.
4.4 Tight-binding parameters, coupling energies

4.4.1 Tight-binding: statement of the problem

The band structure of a quantum cascade laser can be divided into sub-periods. As the scattering limits the extension of the wavefunctions, these divisions are relevant.

The current across a coupling barrier is usually modeled by sequential resonant tunneling. In this model, some parameters like the coupling energy between two states separated by a barrier need to be obtained from the band structure. This energy arises from the matrix element $\langle \psi_L | V | \psi_R \rangle$ where $V$ is the coupling barrier potential the localized states $\psi_L$ and $\psi_R$, belonging to the Left/Right sub-periods, referenced from the coupling barrier. The coupling energy is therefore depends essentially on the width of the barrier. It basically involves the overlap between the wavefunctions. As the wavefunctions decay exponentially in the barrier, the coupling energy will be roughly found to depend exponentially on the barrier width.

We will examine the case of two quantum states coupled by a barrier of fixed width. We will illustrate the calculations on the band-structure of the two-phonon quantum cascade structure examined in (4.3). The calculations are based on the case of two coupled quantum wells discussed in [8]. The calculations are not complicated but a little bit cumbersome, due to the non-parabolicity of the conduction band. Moreover the non-orthonormality of the states implies a careful treatment of the overlap factors.

4.4.2 Calculation of the coupling energies

We consider two quantum states $|\psi_L\rangle$ and $|\psi_R\rangle$ coupled by a barrier of fixed width. The states $|\psi_L\rangle$ and $|\psi_R\rangle$ are eigenstates of their respective hamiltonian with energies $E_L$ and $E_R$. These states are not in general the ground-states of two identical potentials. We even do not require that $E_L = E_R$ as in usual tight-binding models [4]. However we need to keep in mind that this model will hold for a pair of states near the resonance condition and well-separated from the other states. Although we can approximate the general case by treating all resonances independently, the solution remains basically a two-state approximation. The states $|\psi_L\rangle$ and $|\psi_R\rangle$ are solution of:
\[ H_L \psi_L = E_L \psi_L \quad H_L = p_z \frac{1}{2m(z, E_L)} p_z + V_L(z) \]

\[ H_R \psi_R = E_R \psi_R \quad H_R = p_z \frac{1}{2m(z, E_R)} p_z + V_R(z) \]

where \( V_{L,R} \) is the potential for the left/right region. We assume that the states \( \psi_L \) and \( \psi_R \) are bounded to their respective potential region. There exist a point \( Z_L \) after which the potential region \( V_L \) is forbidden for energy \( E_L \): \( V_L - E_L < 0 \) for all \( z > Z_L \). A similar condition holds for the right potential \( V_R \), where \( V_R - E_R < 0 \) for all \( z < Z_R \). With \( Z_L < Z_R \), there exists a barrier potential region between the left and right regions where the coupling energy between the two localized states can be computed.

In the above equations, we have explicitly used the energy-dependent mass computed at the eigen-energies \( E_{L,R} \).

We introduce the total potential \( V(z) \): the potential \( V_L \) in the left region \( ]-\infty, Z_L[ \) clipped

\[ \begin{align*}
V_L(z) & \quad V_C(z) \\
E_L & \quad V_R(z)
\end{align*} \]

Figure 4.5: The tight-binding problem is illustrated. The left localized potential \( V_L(z) \) is shown in green and the right localized potential \( V_R(z) \) in red. The energies \( E_L \) and \( E_R \) are reported. The \( z \)-axis is separated into regions illustrated by shaded areas. The coupling region is outlined in orange between \( Z_L \) and \( Z_R \). The coupling potential \( V_C(z) \) was chosen as the envelop of the localized potentials. This choice is arbitrary. In order to relate the general case to standard cases found in heterostructures, we have added on the illustration a standard configuration (dashed lines) found in heterostructures. The potential (dashed black) is a step-like potential. The localized left/right potentials are quantum wells and the coupling potential is a barrier. The shape of the wavefunctions (dashed green/red) are also represented. The validity of the model is however not limited to ground-states.
at $Z_L$, the potential $V_R$ in the right region $|Z_R, \infty|$ clipped at $Z_R$, and a coupling potential $V_c$ defined on $[Z_L, Z_R]$. We can write formally:

$$V(z) = \theta(Z_L - z)V_L(z) + \theta(z - Z_R)V_R(z) + \theta(z + Z_L)(1 - \theta(Z_R - z))V_c(z) \quad (4.19)$$

The necessary condition for bound-states $V_c(z) > E_{L,R}$ is required. The Fig. (4.5) illustrates the problem.

The potential $V_c$ is a priori arbitrary. In the simplest case it is a flat barrier. However we need to treat the general case, since an arbitrary electro-static potential $q_0 \Phi$ can be added.

In a two-state problem, the coupled wavefunction $\phi$ is a linear superposition of the localized wavefunctions:

$$|\phi\rangle = \alpha |\psi_L\rangle + \beta |\psi_R\rangle \quad (4.20)$$

The parameters $\alpha$ and $\beta$ are complex numbers.

The state $\phi$ is an eigenstate of the complete hamiltonian $\mathcal{H}$ defined for any position $z$ in $\mathbb{R}$:

$$\mathcal{H}|\phi\rangle = E|\phi\rangle \quad \text{with} \quad \mathcal{H} = p_z \frac{1}{2m(z,E)}p_z + V(z). \quad (4.21)$$

Here we implicitly assumed that the energy and position dependent mass $m(z,E)$ is well-defined on the coupling region $[Z_L, Z_R]$. We have evaluated it at the energy of the coupled eigenstate.

We now need to project this equation on the localized states. We introduce the usual notation for the scalar product instead of the usual bra-ket notation. We also drop the ket and bra notations. For example, for vectors $f$ and $g$, and a linear operator $A$, we have: $(f|Ag)$ instead of $\langle f|A|g\rangle$. This notation is useful when the product $fAg$ is ordered.

We project Eq. (4.21) on $\psi_L$ and $\psi_R$, and we insert the decomposition of $\phi$ on localized states. This yields:

$$\langle \psi_L|\mathcal{H}\phi \rangle = \frac{1}{2} \langle \psi_L|p_z m^{-1}(z,E)p_z \phi \rangle + \langle \psi_L|V(z)\phi \rangle$$

$$= \frac{1}{2} \alpha \langle \psi_L|p_z m^{-1}(z,E)p_z \psi_L \rangle + \frac{1}{2} \beta \langle \psi_L|p_z m^{-1}(z,E)p_z \psi_R \rangle + \alpha \langle \psi_L|V(z)\psi_L \rangle + \beta \langle \psi_L|V(z)\psi_R \rangle \quad (4.22)$$

We can identify the kinetic $K_{ij}$ and potential $T_{ij}$ terms; the $T$ notation for the potential terms refers to the transfer integral in usual tight-binding models. The following definitions clarify the notation and help to keep track of the approximations:

$$K_{ij}(E) = \frac{1}{2} \langle \psi_i|p_z m^{-1}(z,E)p_z \psi_j \rangle \quad \text{and} \quad T_{ij} = \langle \psi_i|V(z)\psi_j \rangle. \quad (4.23)$$
For the projections \((\psi_L|\mathcal{H}\phi)\) and \((\psi_R|\mathcal{H}\phi)\), we have:

\[
(\psi_L|\mathcal{H}\phi) = \alpha (K_{LL}(E) + T_{LL}) + \beta (K_{LR}(E) + T_{LR}) \\
(\psi_R|\mathcal{H}\phi) = \alpha (K_{RL}(E) + T_{RL}) + \beta (K_{RR}(E) + T_{RR})
\]  

(4.24)

Using the eigen-equation \(\mathcal{H}\phi = E\phi\), we have:

\[
(\psi_L|\mathcal{H}\phi) = E (\alpha r_{LL} + \beta r_{LR}) \quad \text{and} \quad (\psi_R|\mathcal{H}\phi) = E (\alpha r_{RL} + \beta r_{RR})
\]  

(4.25)

where the overlap integrals are defined by \(r_{ij} = (\psi_i|\psi_j)\). We use the notation \(r_j \equiv r_{jj}\).

Since \(r_{ij} = r_{ji}\), we set \(r_{RL} = r_{LR} = r\). The system given by Eq.(4.24) and Eq.(4.25) reads in matrix form:

\[
H(E) \cdot v = E R \cdot v,
\]

(4.26)

The \(H\) and \(R\) matrices are given by:

\[
H(E) = \begin{pmatrix} K_{LL}(E) + T_{LL} & K_{LR}(E) + T_{LR} \\ K_{RL}(E) + T_{RL} & K_{RR}(E) + T_{RR} \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} r_{LL} & r \\ r & r_{RR} \end{pmatrix}
\]  

(4.27)

The operator \(H(E)\) depends on the eigenvalue \(E\). We need to do some simplifications. The energy dependence is given by the kinetic term \(K(E)\). We remove this dependence by evaluating \(K\) at the energy of the localized wavefunction. Without introducing a new notation, we have for localized state \(\psi_j\) with energy \(E_j\):

\[
K_{ij} = \frac{1}{2} (\psi_i|p_z m^{-1}(z, E_j)p_z \psi_j)
\]  

(4.28)

The eigen-equation for the coupled states can be rewritten is the usual form \(\tilde{H} \cdot v = E v\) by multiplying the equation by the matrix \(R^{-1}\). We simplify the notations by defining: \(H_i = K_{ii} + T_{ii}\) and \(H_{ij} = K_{ij} + T_{ij}\). The matrix elements of \(\tilde{H} = R^{-1}H = (h_{ij})\) are:

\[
\begin{align*}
    h_{LL} &= \delta^{-1}(r_{R}H_{L} - r_{H_{LR}}) \\
    h_{LR} &= \delta^{-1}(r_{R}H_{LR} - r_{H_{R}}) \\
    h_{RL} &= \delta^{-1}(r_{L}H_{RL} - r_{H_{L}}) \\
    h_{RR} &= \delta^{-1}(r_{L}H_{R} - r_{H_{LR}}).
\end{align*}
\]  

(4.29)

The determinant \(\delta = r_{LR}r_{R} - r^2\) comes from the inversion of matrix \(R\).

The diagonalization of \((h_{ij})\) yields the solutions:

\[
E_\pm = -\frac{h_{LL} + h_{RR}}{2} \pm \sqrt{\left(\frac{h_{LL} - h_{RR}}{2}\right)^2 - h_{LR}h_{RL}}
\]  

(4.30)
4.4. Tight-binding parameters, coupling energies

The latter can be directly related to a standard two-state problem [20] where:

\[
H_0 + H_C = \begin{pmatrix}
E_1 + S_1 & h\Omega \\
h\Omega & E_2 + S_2
\end{pmatrix}
\]  

(4.31)

The unperturbed hamiltonian is given by \( H_0 = \text{diag}(E_1, E_2) \). The \( S_{1,2} \) matrix elements are shift-energies: they don’t couple the unperturbed states but simply shift the eigen-energies. The element \( h\Omega \) is the coupling energy.

The comparison with the tight-binding hamiltonian gives the identification:

\[
h^2\Omega^2 = h_{LR} h_{RL} \quad \text{and} \quad \Delta = h_{LL} - h_{RR}.
\]  

(4.32)

where the renormalized detuning \( \Delta \) between shifted energies has been introduced. It is in general not equal to the detuning between the localized states \( \Delta_0 = E_L - E_R \). It is worth to mention that we have used a special normalization \( r_L \) and \( r_R \), as in the effective two-band model (3.33). But this care is not necessary for computing the coupling energy as the latter is invariant under renormalization of the localized states. This can be checked with the transformations \( \psi_L \to c\psi_L \) and \( \psi_R \to d\psi_R \), where \( c \) and \( d \) are constants.

Even if the \((h_{ij})\) matrix is not symmetric, the coupling energy is well-defined.\(^4\)

4.4.3 Evaluation of \( K_{ij} \) and \( T_{ij} \)

We now evaluate the kinetic \( K_{ij} \) and potential \( T_{ij} \) terms for each combination of \( L \) and \( R \) indices. We will need to make an approximation on the computation of the effective mass. The latter is not necessary but simplify the computation. We now work in the position representation with wavefunctions \( \psi_L(z) \) and \( \psi_R(z) \). For bound-states we can choose real wavefunctions. The potential term \( T_{ij} \) is given by:

\[
T_{ij} = (\psi_i|V|\psi_j) = \int dz \psi_i(z)V(z)\psi_j(z).
\]  

(4.33)

All terms under the integral are well-defined. This is not the case for the kinetic term. By using the momentum \( p_z = -i\hbar \partial_z \), we have:

\[
2K_{ij} = (\psi_i|p_z m^{-1}(z, E_j)p_z \psi_j) = \int dz \psi_i(z)p_z m^{-1}(z, E_j)p_z \psi_j(z) = -\hbar^2 \int dz \psi_i(z)\partial_z (m^{-1}(z, E_j)\partial_z \psi_j(z)).
\]  

(4.34)

\(^4\)In a numerical implementation it is safer to set \( \hbar\Omega = |h_{LR} h_{RL}| \) as small coupling energies can be found negative (numerical artifacts).
We have abusively used the partial-differentiation notation for \( z \)-only dependent functions in order to compact the notation. We evaluate the following integral:

\[
\int dz \frac{d}{dz} \left( \psi_i(z) \frac{1}{m(z, E_j)} \frac{d}{dz} \psi_j(z) \right) = \int dz \left( \frac{d}{dz} \psi_i(z) \right) \frac{1}{m(z, E_j)} \left( \frac{d}{dz} \psi_j(z) \right)
\]

\[+ \int dz \psi_i(z) \frac{d}{dz} \left( \frac{1}{m(z, E_j)} \frac{d}{dz} \psi_j(z) \right) \tag{4.35}\]

and:

\[
\int dz \frac{d}{dz} \left( \psi_i(z) \frac{1}{m(z, E_j)} \frac{d}{dz} \psi_j(z) \right) = \psi_i(z) \frac{1}{m(z, E_j)} \frac{d}{dz} \psi_j(z) \bigg|_{-\infty}^{\infty} \tag{4.36}\]

\[= 0 \]

since \( \psi_i(z) \to 0 \) when \( |z| \to \infty \) (bound-stats). This yields the identity:

\[
\int dz \psi_i(z) \frac{d}{dz} \left( \frac{1}{m(z, E_j)} \frac{d}{dz} \psi_j(z) \right) = - \int dz \left( \frac{d}{dz} \psi_i(z) \right) \frac{1}{m(z, E_j)} \left( \frac{d}{dz} \psi_j(z) \right). \tag{4.37}\]

The kinetic term is therefore expressed as:

\[
K_{ij} = \frac{\hbar^2}{2} \int dz \left( \frac{d}{dz} \psi_i(z) \right) \frac{1}{m(z, E_j)} \left( \frac{d}{dz} \psi_j(z) \right) \tag{4.38}\]

The wavefunctions derivatives are not continuous in a layered medium with a position-dependent effective mass. The first continuous derivative is: \( D\psi_i(z) = m^{-1}(z, E_i) \partial \psi_i(z) \). However, the mass considered for the kinetic term, is the mass of the coupled system while the mass entering the definition of \( D\psi_i \) is the mass of the localized system. We therefore use the notation \( D_0 \psi_i \) to underline the use of the mass of the localized system. As \( D_0 \psi_i \) is already known it is practical to use it:

\[
K_{ij} = \frac{\hbar^2}{2} \int dz \left( \frac{d}{dz} \psi_i(z) \right) \frac{1}{m(z, E_j)} \frac{1}{m(z, E_j)} \left( \frac{d}{dz} \psi_j(z) \right) \tag{4.39}\]

In the last step, we have made the approximation that the mass of the localized system coincide with the mass of the coupled system \( m_j(z, E_j) = m(z, E_j) \). We have extended
this equality for the state $i$: $m_i(z, E_i) = m(z, E_i)$. For state $\psi_L$, for example, this is true for $z \in ]-\infty, Z_L[$ but in general not in the coupling barrier region $[Z_L, Z_R]$ and not in the right region $]Z_R, \infty[$. However in quantum cascade lasers, the left localized potential ends with an infinite barrier, while in the coupling region it is replaced by a finite length coupling barrier. The masses therefore coincide in the coupling barrier region. In the right region this approximation does not hold, but if the barrier is thick compared to the penetration length of the wavefunctions, the error is negligible. This is of course not be the case for thin barriers.

The mass is evaluated at energy $E_i$ in the approximated result in Eq.(4.39). This approximation has given accurate results in transport simulations and it can therefore be used directly in numerical implementations. A numerical issue was the extension of the finite region of left/right potentials and masses to the real line. The effective mass of the left/right region is in general ill-defined numerically. This was the basic reason for this approximation, although it is questionable for thin coupling barriers, as discussed above.

Here we summarize the expressions $K_{ij}$ and $T_{ij}$:

$$K_{ij} = \frac{\hbar^2}{2} \int dz \, D_0 \psi_i(z) m(z, E_i) D_0 \psi_j(z) \quad T_{ij} = \int dz \psi_i(z) V(z) \psi_j(z).$$

(4.40)

If we use the propagation algorithm (4.2) in order to solve the localized potentials $V_L(z)$ and $V_R(z)$ on a finite region, thanks to the wavefunction coefficients $(A_i, B_i)$, we know the wavefunctions $\psi_{L,R}(z)$ and its first continuous derivative $D_0 \psi_{L,R}(z)$ for $z \in \mathbb{R}$. Giving the exact form (typically a barrier) for the coupling potential $V_C(z)$ for $z \in [Z_L, Z_R]$, we can compute $T_{ij}$ directly for $i, j \in \{L, R\}$. The effective mass $m(z, E)$ needs to be defined on $[Z_L, Z_R]$. Then we can compute $K_{ij}$. After having computed the overlap matrix $R$, the coupling energy $h\Omega$ can be directly obtained with Eq.(4.32) and Eq.(4.29).

For localized bases $\{\psi^{(i)}_L\}$ (dimension N) and $\{\psi^{(j)}_L\}$ (dimension M), we can compute the two-state coupling parameters for all possible pair of states. This yields a coupling matrix $M_\Omega = (h\Omega_{IJ})$ where $I = 1..N$ runs in the left basis and $J = 1..M$ runs in the right basis. It is worth to mention that we do not have solved the more general problem of $N$ left-localized states coupled to $M$ right-localized states. Although the coupling matrix $M_\Omega$ can be mapped onto a well-defined Hamiltonian, the shift energies $h_{LL}$ and $h_{RR}$ giving a renormalized detuning $\Delta$ for each pair of states cannot be represented by a $(N + M) \times (N + M)$ matrix. These energies affect only the diagonal elements of the Hamilton matrix. We therefore have an over-determined system with $2NM$ terms for $N + M$ independent elements. Renormalized detuning are therefore only available for the general solution.

The coupling matrix $(M_\Omega)_{IJ}$ is not symmetric. Its left index runs over left localized
states, while its right index runs over right localized states. It will be represented by a $(N + M) \times (N + M)$ Hamiltonian when the coupling between sub-periods in quantum cascade structures will be treated (7).
4.4.4 Numerical examples

Simplest case: coupled quantum wells

We consider a numerical example: the case of two quantum wells of width $L$ coupled by a barrier of width $W$ [8]. The wells/barriers are made in Ga$_{0.47}$In$_{0.53}$As /Al$_{0.48}$In$_{0.52}$As lattice matched on InP. The width of the well is 3.0 nm. The single quantum well have only one bound-state at energy 220 meV (with infinite barriers on both sides). Although the numerical technique we have developed is optimized for computing the coupling energies in quantum cascade structures, we can still use it to compute the coupling between ground-states of coupled quantum-wells. However we have to be careful by choosing a large enough barrier: the penetration length of the wavefunctions should not exceed the barrier width. Else, we are running out of the previous approximations.

![Coupling energy vs Barrier width](image)

Figure 4.6: The coupling energy between the ground-states of two coupled quantum-wells is shown in logarithmic scale against the coupling barrier width. The material system is Ga$_{0.47}$In$_{0.53}$As /Al$_{0.48}$In$_{0.52}$As lattice matched on InP. The solid line represents the tight-binding model approximation, while the dashed line reports the half of the detuning-energy between the coupled states, when the two-well potential is solved at once. The band-structure is reported for two barrier widths: 8.0 nm and 1.5 nm. In the shaded area the approximations do not hold.
In Fig. (4.6) the computed coupling energy is reported against the barrier width. The barrier is varied from 1 Å to 100 Å. The dependence of the coupling energy is clearly exponential with respect to the barrier width. If we push the model to its limit the coupling energy remains well-defined. The exponential region between 20 Å and 80 Å has been fitted with an exponential with a characteristic length of 15.2 Å. This number gives the typical barrier width below which the model is no more accurate. This number is however related to the single quantum-well region. For quantum cascade structures the validity range can be different.

In Fig. (4.6) we show the results given by the tight-binding approximation (solid line) and the exact results (dashed line), where the coupling energy is given by $\hbar \Omega = \hbar \Delta / 2$, the detuning-energy $\hbar \Delta$ being evaluated between the coupled-states when the whole potential is solved at once. The tight-binding model predicts about 80% of the expected coupling energy in a two-state system. For a first-order approximation this however a good result. We have nevertheless checked if it was not related to a normalization issue. It is not the case, as the coupling energy obtained from the tight-binding model is invariant under renormalization of the wavefunctions.

**Coupling energies in two-phonon structure**

We have applied the tight-binding model to the two-phonon structure presented in (4.3.2) at the alignment field. We use the notation in Fig. (4.4). We first evaluate the coupling between the injector ground-state G and the upper-laser level 4. We have: $\hbar \Omega_{G4} \approx 2.53$ meV. The extraction resonances coupling energies are: $\hbar \Omega_{3A} \approx 5.0$ meV, $\hbar \Omega_{2B} \approx 4.0$ meV and $\hbar \Omega_{1C} \approx 1.9$ meV.

In a transport model we need to account for all (other) resonances. However the current is mainly driven by these ones.
4.5 Self-consistent potential

In solid-state physics the many-body problem cannot be solved, we therefore develop averaged one-particle potentials [23]. We already have performed calculations in this approximation for solving the band-structure of semiconductors in (2.4). The simplest approximation of the electron-electron approximation is called the Hartree approximation [8]. In the one-electron problem the particle experiments a self-consistent potential due to the other electrons. The latter arises from the distribution of the electrons inside the structure. The problem therefore consists of coupled Schrödinger and Poisson equations. We have neglected here the exchange energy due to the anti-symmetry of the total wavefunction under the exchange of two electrons (Pauli principle). If the exchange energy is taken into account this yields the Hartree-Fock model.

This part of the work is found between the band-structure (as the carrier distribution will bend the potential) and the transport modeling, since we need to know the carrier distribution.

We assume that the carriers distribution is known for states \( \{ \phi_i \} \) with concentrations \( n_i \). As we work in the envelop function approximation, we do not make assumptions over the intra-subband distribution of the carriers. The two-dimensionality of the subbands will be relevant when we will compute the Fermi-level for a thermal distribution.

When a charge distribution is assumed, the potential is called self-consistent, since the charge distribution computed for a particular band-structure (by a transport model, or a thermal distribution) will alter the potential and therefore the eigen-energies. In this new basis, the charge distribution is recomputed, therefore altering again the potential. It is clear that the computation of the self-consistent potential requires iterations and a convergence criterion. In this section, we do not pay attention to the calculations of the densities from the band-structure. The convergence technique is briefly mentioned at the end of this section.

We have not discussed so far the origin from the charge density in a heterostructure. A low electron-concentration is given in the conduction band by doping. The usual dopant is Si. The doped layers (well and barriers) are chosen so that the dopants are far as possible from the region where the laser transition occurs. If the dopants were implanted at the same position than the laser-doublet, on the one hand the upper laser-state lifetime would be strongly reduced due to inter-subband transitions caused by the dopants (5.3.3), and on the other hand the gain optical linewidth (5.2.3) would be largely increased by both inter-subband scattering and dephasing rates arising from intra-subband mechanisms (5.1.2). Even for moderate volume concentrations, the population inversion would be destroyed.
and the optical linewidth several times larger.

The doped layers are therefore chosen in the injector region. In Fig.(4.4), the pink-shaded region in the injector is doped with a volume concentration of electrons of $0.2 \cdot 10^{18} \text{ cm}^{-3}$. This gives a sheet carrier density $n_s = 2 \cdot 10^{11} \text{ cm}^{-2}$ by multiplying the volume concentration by the thickness of the doped region. The doping is a very subtle issue in heterostructures and especially for quantum cascade lasers - the background doping can alter very significantly the operation of the structures [1]. We will not discuss these problems here and we only consider that layers are nominally doped to a certain volume concentration, providing a sheet carrier density per period of the quantum cascade structure. The position of the dopants in the structure will also be important in the evaluation of impurity-scattering (5.3.3).

4.5.1 Modeling the self-consistent potential

We assume that a multi-quantum-well structure defined on $[0, L]$ accommodates $M$ bound-states $| \phi_i \rangle$ with $i = 1..M$. These bound-states are populated with a concentration of $n_s$ electrons. If $n_i$ is the concentration for bound-state $\phi_i(z)$, we impose the charge conservation with $\sum_i n_i = n_s$. The charge density associated with the distribution is:

$$\rho_e(z) = -\sum_{i=1}^{M} |\phi_i(z)|^2 n_i \quad \text{with} \quad \langle \phi_i | \phi_i \rangle = 1 \quad \text{for} \quad i = 1..M.$$  

(4.41)

The charge density is here normalized with an elementary charge of 1. The sign-less elementary charge $q_0$ will multiply the corresponding electro-static potential: $q_0 \Phi(z)$. We have explicitly assumed that the wavefunctions are normalized to unity. While the propagation coefficients for the wavefunctions $\phi_i(z)$ are known on $[0, L]$, the wavefunctions are defined on the whole real line. This requirement is not necessary in a numerical implementation, but it gives in a first attempt a simpler formulation. We therefore have:

$$\int_{\mathbb{R}} dz \rho_e(z) = -n_s.$$  

For $\rho_e(z)$ only the structure is not electrically neutral. We need to consider the charge density $\rho_d(z)$ due to the ionized dopants. The latter is essentially a step function in the doped layer, but we simply require that $\int_{\mathbb{R}} \rho_d(z) = n_s$. The charge neutrality is therefore ensured:

$$\int_{\mathbb{R}} dz (\rho_e(z) + \rho_d(z)) = 0.$$  

Although the charge neutrality is ensured at the boundary of the real line, the local net charge density $\rho(z) = \rho_e(z) + \rho_d(z)$ gives rise to an electro-static potential $q_0 \Phi(z)$. The latter is calculated by solving the Poisson equation in a dielectric medium with constant $\epsilon$: 

\[ \nabla^2 \Phi = -\rho/\varepsilon_0 \varepsilon. \]

For a one dimensional problem in a layered dielectric medium, we have:

\[ \frac{d^2}{dz^2} \Phi(z) = -\frac{\rho(z)}{\varepsilon_0 \varepsilon(z)}. \]  (4.42)

We integrate this equation with initial conditions at \( z = z_0 \), for the potential and the electric field \( E = -q_0 \nabla \Phi \). This yields:

\[ \Phi(z) = \Phi(z_0) + \frac{\mathcal{E}(z_0)}{q_0} (z - z_0) - \int_{z_0}^{z} dx \int_{z_0}^{x} d\tau \frac{\rho(\tau)}{\varepsilon_0 \varepsilon(\tau)}. \]  (4.43)

We have to set the boundary conditions \( \phi(z_0) \) and \( \mathcal{E}(z_0) \) carefully. We assume that the quantum cascade period is defined on interval \([0, L_p] \). We consider the self-consistent contribution only. An external applied (constant and uniform) electric field will appear as an additional electro-static contribution \( -q_0 F z \) in the hamiltonian. We have to ensure that the value of the potential at one boundary of the period \( \Phi(0) \) is equal to the value at the other boundary of the period \( \Phi(L_p) \). Therefore \( \Delta \Phi = \Phi(L_p) - \Phi(0) \equiv 0 \): the self-consistent potential does not change the potential at the boundary of the period. In the above general solution, we set \( z_0 = 0 \) and we compute \( \Phi(L_p) \). This gives a condition on \( \mathcal{E}(0) \):

\[ \mathcal{E}(0) = \frac{q_0}{L_p} \int_{0}^{L_p} dx \int_{0}^{x} d\tau \frac{\rho(\tau)}{\varepsilon_0 \varepsilon(\tau)}. \]  (4.44)

We remove the constant offset from the electro-static potential by setting \( \Phi(0) = 0 \). Finally the self-consistent potential is defined on \( z \in [0, L_p] \):

\[ \Phi(z) = \frac{\mathcal{E}(0)}{q_0} z - \int_{0}^{z} dx \int_{0}^{x} d\tau \frac{\rho(\tau)}{\varepsilon_0 \varepsilon(\tau)}. \]  (4.45)

This expression is compatible with periodic boundary conditions. The dielectric constant \( \varepsilon(z) \) is a position-dependent function as different materials are used for well and barriers.

There are different ways of solving the Poisson equation and obtaining \( \Phi(z) \) with the right boundary conditions. We consider one period \([0, L_p] \). We impose the periodicity by expanding \( \Phi(z) \) in a Fourier serie. Let \( f(z) \) be a periodic function with fundamental domain \([0, L] \). Its Fourier expansion is given by:

\[ f(z) = \sum_{n} c_n \exp \left( i \frac{2\pi}{L} x \right) \quad \text{with} \quad c_n = \frac{1}{L} \int_{0}^{L} dx \ f(x) \ \exp \left( -i \frac{2\pi}{L} x \right). \]  (4.46)

The potential \( \Phi(z) \) is then represented by the Fourier coefficients \( \varphi_n \). We have to solve \( \partial^2 \Phi = F(z) \), where \( F(z) \) is a periodic function on \([0, L] \) with Fourier coefficients \( f_n \). If we
insert the expansion of \( \Phi \) and \( F \) in the Poisson equation we get an algebraic equation for the coefficients:

\[
\varphi_n = - \left( \frac{L}{2\pi} \right)^2 \frac{1}{n^2} f_n \quad \text{for } n \text{ integer, but not zero.} \tag{4.47}
\]

The constant \( \varphi_0 \) for \( n = 0 \) is free. If now we apply the boundary conditions for the periodic heterostructure, we set the constant offset to zero: \( \varphi_0 = 0 \). The periodicity of the solution is accounted by definition. We have:

\[
\Phi(z) = - \left( \frac{L}{2\pi} \right)^2 \sum_{n \neq 0} \frac{f_n}{n^2} \exp \left( \frac{2\pi}{L} x \right). \tag{4.48}
\]

If \( F(z) = -\rho(z)/\varepsilon_0\varepsilon(z) \) is chosen.

### 4.5.2 Periodicity of the charge density

This last equality rises an important question. When we consider the net charge density \( \rho(z) \), the electron concentration of each state is weight by the square modulus of its wavefunction as in Eq.(4.41). The wavefunctions are not periodic since they are solutions of an insulated single period. We therefore need to consider periodicity of the probability density before we solve the self-consistent potential.

The wavefunctions \( \phi_i(z) \) are not periodic with the length of the period \( L_p \), but they decay with increasing \( |z| \) and are zero at infinity. Here we will force the periodicity by extending the wavefunctions computed for a single insulated period to neighboring periods. It is worth to mention that we do not solve the potential of three adjacent periods but that we extend the wavefunction by extrapolation (with last propagation coefficients) to adjacent periods. Usually we do not need to use more than one adjacent period, as the wavefunctions are exponential at the boundary of the fundamental period.

We then fold the lateral periods into the fundamental period. The left period \([-L_p, 0] \) is shifted on the fundamental period \([0, L_p] \), while the right period \([L_p, 2L_p] \) is shifted back to \([0, L_p] \). For the wavefunctions we have to sum the squared moduli. For the density of probability \( |\phi_i|^2(z) \), we define its value on the left period by \( |\phi_i^{(-1)}|^2(z) = |\phi_i|^2(z + L_p) \) and its value on the right period by \( |\phi_i^{(+1)}|^2(z) = |\phi_i|^2(z - L_p) \). The periodic squared moduli (truncated to the nearest period) \( |\overline{\phi_i}|^2 \) is then:

\[
|\overline{\phi_i}|^2 = |\phi_i|^2 + |\phi_i^{(-1)}|^2 + |\phi_i^{(+1)}|^2.
\]

In general we can write:

\[
|\overline{\phi_i}|^2(z) = |\phi_i(z)|^2 + \sum_n \left( |\phi_i^{(-n)}|^2(z) + |\phi_i^{(+n)}|^2(z) \right). \tag{4.49}
\]
4.5. Self-consistent potential

However the nearest neighboring period is largely sufficient. The periodic density $|\phi_i|^2(z)$ is normed to unity on $[0, L_p]$ if $\phi_i$ was normed to unity on the real line.

The electron charge density $\rho_e(z)$ in Eq.(4.41) has to be constructed with the periodic densities $|\phi_i|^2(z)$. The charge density is then well-defined on the fundamental domain $[0, L_p]$. The electro-static potential can be obtained by direct integration or by Fourier expansion as discussed above.

4.5.3 Example of self-consistent potential

We give an example of structure where the Hartree self-consistent potential is quantitatively important. The structure itself is not very interesting but it shows a typical situation where the self-consistent potential cannot be neglected in the band-structure computation.

The Eq.(4.45) shows that the potential is generated by the net local charge density. It implies that the self-consistent potential will become more important as the distance between electrons and the dopants increases. The structure we consider is shown in Fig.(4.7). It is designed to have the ground-state at the left of the structure and the dopant at the right of the structure. We have used a high sheet carrier density of $n_s = 3 \cdot 10^{11}$ cm$^{-2}$ concentrated in a thin well of 3.0 nm in order to maximize the self-consistent potential. All the electrons are put in the ground-state (representing a thermal distribution at low temperatures, about 10 K). The self-consistent potential is solved by iteration and the convergence on the eigen-energies is briefly explained in (4.5.6). The contribution of the self-consistent potential is clearly visible in Fig.(4.7b) as the multi-quantum-well potential shown in the part (a) of the figure is bent. The self-contribution alter seriously the spectrum of the structure, as the ground-state energy varies from 103.3 meV to 134.1 meV. It is also the case for other eigen-states: the spectrum is "compressed".

The transport can be strongly influenced by the self-consistent potential, as eigen-energies and therefore scattering and resonances are modified. In mid-infrared quantum cascade lasers, this effect can often be neglected. However in far-infrared structures (typically grown in AlGaAs/GaAs on GaAs substrate) the effect of the self-consistent potential is crucial in the understanding of the operation of the structure (6.5).

For the test structure presented here, we have chosen a very thick coupling barrier of 10.0 nm. This choice enables the approximation $|\phi_i|^2(z) \approx |\phi_i|^2(z)$. But in general we have to force the periodicity in the fundamental period in order to model accurately the effect of the neighboring periods on the coupling barrier.
Figure 4.7: A simple structure designed to enhance the self-consistent potential is shown. The structure is made of Ga$_{0.47}$In$_{0.53}$As /Al$_{0.48}$In$_{0.52}$As lattice matched on InP. The layer sequence starting from the first barrier at the left of the structure is in nm: 10.0, 6.0, 4.0, 5.0, 4.0, 4.0, 4.0, 3.0. The last well, as indicated by the shaded pink area is volume doped to $1 \cdot 10^{18}$ cm$^{-3}$, giving a sheet carrier density of $n_s = 3 \cdot 10^{11}$ cm$^{-2}$. (a) The band-structure is shown without self-consistent potential. The whole electron concentration is put in the ground-state (in red). Its energy without the self-consistent potential is 103.3 meV. (b) The band-structure is solved with the self-consistent potential. The energy of the ground-state is 134.1 meV. The potential region is bent by the local self-contribution. The dashed lines show that the net contribution of the self-consistent potential is zero across the period. The last barrier at the right is a repetition of the first barrier, implementing periodicity of the potential.

4.5.4 Modeling issues

In the previous paragraphs we have solved the self-consistent potential for one period of the heterostructure with periodic boundary conditions. We have assumed that the eigen-states are computed for one-period. As shown in Fig.(4.3a) the eigen-energies are determined by propagation of the solution across the whole period. However as shown in Fig.(4.4), structures like the two-phonon design, need to be split at an exit barrier inside the period. This is required when we consider coherent effects at several coupling barriers. In this latter case, we cannot use the eigen-states computed for the whole period in order to determine the self-consistent potential.
In order to get the localized wavefunctions, we solve each sub-period independently (for example, the active region and the injector region for the two-phonon design). We get localized bases \( \{ \phi_i(z) \} \) with \( I = 1..M \) for \( M \) sub-periods and \( i = 1..N_I \) for each eigen-basis of dimension \( N_I \). We then reconstruct the complete period by juxtaposing each sub-period with the corresponding coupling barrier (at which the period was cut). We also shift the localized wavefunctions to put them at the right place inside the reconstructed potential. With this procedure, we obtain the potential and the wavefunctions shown in Fig.(4.4).

We then force the periodicity of each wavefunction of this joined basis. It is then assumed that a transport model can be solved and gives the population in each localized state. This enable us to define the charge density and therefore the self-consistent potential for the whole period even in the case of localized sub-periods.

The technique described above is used in a first step, when we compute the self-consistent from the bare potential. However this technique can be applied in the same way when we need to do more iterations in order to converge on a solution with a target accuracy on the eigen-energies. The self-consistent potential found in the previous iterations is added to the potential of the sub-periods and the eigen-bases are computed. This enables us to get a new self-consistent potential. We add it to the unbent potential and the procedure described in the above paragraph is the same. We have underlined that at each step we obtain a new self-consistent potential. This gives a series of self-consistent potentials: \( \Phi_0, \Phi_1, .. \). But we may think in terms of variation of the self-consistent potential with \( \delta \Phi_1 = \Phi_1 - \Phi_0 \), etc. This representation is sometimes more intuitive.

In (4.5.6) we discuss the convergence algorithm used for the self-consistent potential. We treat this problem in the current section in order to avoid to have to postpone it when we discuss the transport models, as it is useless to clutter the discussion with these methods.

### 4.5.5 Self-consistent potential from a thermal distribution

The self-consistent potential can be computed for any transport model, as we only need as input the wavefunctions, the populations and the doping profile only. However this can be heavy in terms of computational effort since the transport model needs to be solved at each iteration until convergence.

In some case, it is more useful to assume that the carriers, that mainly reside in the injector region can be described with a good accuracy by a thermal distribution in the injector or even in the whole period. This model makes important assumptions on the charge transport in the structure. The first requirement is that the injection barrier defines
clearly the period in terms of charge distribution. In other words, we should be able to define a fermi-level for each period. The main charge should therefore be located in the lower-states of the injector. This is achieved by a selective injection in the upper laser-level (that injects about 10 % of the total charge in the active region) and efficient LO-phonon thermalization in the injector region. As the injector region is typically designed in order to fulfill these requirements, a thermal distribution of carriers is an accurate approximation. As the doped layers and the charge (in a transport regime) are located in the same region, the deviation to the thermal self-consistent potential due to 10 % of the charge in the active region can be neglected in a first approximation. This statement holds typically for mid-infrared structures. For far-infrared structures it is better to solve the self-consistent potential with the populations computed by a transport model.

Although the population in the active region obtained by a thermal distribution in the injector is numerically wrong, the impact on the self-consistent potential is weak as discussed above.

We consider a potential region (basically a period of a quantum cascade structure) that accommodates a basis of bound-states \( \{ \phi_i \} \) with \( i = 1..M \) with eigen-energies \( E_i \) and effective masses \( m_i \) (the definition of the mass incorporate the mass of the bare electron \( m_0 \)). In the envelop function approximation, the averaged effective mass is given by the relation: \( m_i^{-1} = \int dz |\phi_i(z)|^2 m^{-1}(z, E_i) \). At a given thermal energy \( \beta^{-1} = k_b T \), where \( T \) is the temperature and \( k_b \) the Boltzmann’s constant, the total surface concentration of carriers for a Fermi energy \( \mu \) is given by:

\[
N(\mu) = \sum_{i=1}^{M} n_i(\mu) \quad \text{with} \quad n_i(\mu) = \frac{D_0 m_i}{\beta} \log \left( 1 + \exp \left( \beta (\mu - E_i) \right) \right) \quad (4.50)
\]

With \( n_i(\mu) \) being the concentration of carriers in eigen-state \( i \) and \( D_0 m_i \) the density of states for a two-dimensional system of mass \( m_i \), where \( D_0 = (\pi \hbar^2)^{-1} \). The concentration \( n_i(\mu) \) is obtained by integration of the Fermi-Dirac distribution of carriers \( f_i(E, \mu) \) in state \( i \):

\[
f_i(E, \mu) = \frac{1}{1 + \exp \left( \beta (E_i + E - \mu) \right)} \quad \text{with} \quad n_i(\mu) = D_0 m_i \int_0^\infty dE f_i(E, \mu). \quad (4.51)
\]

The above equations hold for a finite temperature \( T > 0 \). The degenerated case \( (T = 0) \) is modeled by:

\[
f_i(E, \mu) = \theta(\mu - (E_i + E)) \quad \text{thus} \quad n_i(\mu) = D_0 m_i (\mu - E_i). \quad (4.52)
\]

For the degenerated case, the Eq.(4.50) needs to be modified according to Eq.(4.52). The quasi-fermi level for each subband \( i \) is defined by \( \mu_i = \mu + E_i \).
4.5. Self-consistent potential

The Fermi-level $\mu_0$ solves $N(\mu_0) = n_s$. This cannot be done algebraically and we need to search numerically the zero of the $N(\mu) - n_s$ function.

We now consider the states bounded to one period of the two-phonon design as shown in Fig.(4.4). We have evaluated the thermal distribution of carriers in the period biased at the alignment field for various temperatures as shown in Fig.(4.8). The self-consistent contribution is not evaluated. As shown in the part (a) of the figure, the electrons are mainly concentrated in a few injector states even at room temperature. We therefore expect that the self-potential is accurately modeled by a thermal distribution. The global Fermi energy shows a weak dependence with the temperature. It remains above the ground-state energy until a (high) temperature of 120 K.

The temperature used to computed the thermal distribution has to be defined more precisely. It is the temperature of the electrons and it cannot be identified with the temper-

![Graph showing electron concentration vs energy for different temperatures.](image)

Figure 4.8: Thermal distribution of the electrons in one period of the two-phonon design is considered. The energies taken for computing the fermi energy $\mu$ are those from the flat potential shown in (b) without the self-consistent potential. (a) the electron concentration in logarithmic scale is reported against the energy in the conduction band. The energies of the injector ground-state, the upper and lower laser-states are marked. The thermal distribution is evaluated for a range of temperatures in Kevlins: 4.2, 15, 77, 150, 320. (b) The band-structure is shown against the energy axis of (a). (c) The Fermi energy is plotted against the temperature.
ature of the lattice (phonon bath). The LO-phonon interaction tends to thermalize the electrons with the lattice. While elastic interactions give in average an excess of kinetic energy in the subbands. The temperature of electrons therefore depends on a detailed balance of the kinetic energy [40]. The latter will be discussed later in this work (7.2). Practically the electrons cannot be cooled to a temperature lower than 100 K. Thus the computations in Fig.(4.4) for temperatures lower than 100 K have no physical meaning for a mid-infrared quantum cascade laser in a transport regime.

Figure 4.9: The self-consistent potential computed for a thermal distribution of carriers with an electronic temperature of 150 K is shown against the spatial position in the band-structure of the two-phonon structure. Between 40.0 nm and 64.0 nm, the density of charge is not zero locally since the electrons (in the injector ground-state) do not overlap with the dopants (shaded area in the band-structure). This causes a small variation of the potential of 12 meV. The higher excited-states in the active region are not shown since they are useless.
In Fig.(4.9) we have performed the calculation of the self-consistent potential for a thermal distribution at 150 K at the alignment field. The self-consistent potential has a very weak influence on the band-structure. It may however cause a renormalization of the injection resonance, but it is practically useless.

This will however not the case for all structures. In a first approximation, the self-consistent potential should be computed for all the band-structures using a thermal distribution. In special designs [99] the evaluation of the self-consistent potential given by a reliable transport model is however crucial. An example is examined in (A).

In this work, we have implemented numerically the computation of the self-consistent potential in virtually all possible cases. The electron concentration can be fixed to a given value in each state, computed by a thermal models, etc.

4.5.6 Self-consistent solution and convergence

We discuss the technique used for the computation of the self-consistent potential to a desired accuracy. The computation is iterative as expected from a self-contribution. A convergence criterion is evaluated at each iteration until the desired accuracy (or a maximal number of iterations) is reached.

The algorithm uses a damping on the populations and on the energies in order to converge. This avoid instabilities that typically arise when the computation is performed at low temperatures. In these cases, the algorithm jumps between two configurations, where the electron are spread over a few states (Fermi energy and ground-state energy are close), or all the electrons are in the ground-state (Fermi energy is far above the ground-state energy).

The algorithm is described below, step by step:

- **(initial)** Initially, the self-consistent potential $\Phi_{(0)}$ is zero.

- **(j.1)** We compute the total potential $V_{\text{tot}}^{(j)}(z) = V(z) - q_0 F z + q_0 \Phi_{(j)}(z)$ in the fundamental period. This is the potential of the heterostructure $V(z)$, an applied constant and uniform electric field $F$ and the self-consistent potential $\Phi_{(j)}(z)$ of the current iteration. We have to consider two cases:

  - The potential $V_{\text{tot}}^{(j)}$ represents a single period. In this case it is directly solved for bound-states, giving a basis $B_1^{(j)}$.

  - The potential $V_{\text{tot}}^{(j)}$ represents a collection of sub-periods that are coupled with several barriers (as in Fig.(4.4) for the exit barrier). The potential is split into
confined sub-period potentials $V^{(j)}_I$. Each of them gives a localized basis $B^{(j)}_I$ with $I = 1..M$ for $M$ sub-periods.

- (j.2) The electrons are distributed over the states of basis $B^{(j)}_I$. The resulting concentrations are $n^{(j)}_{I,i}$ where $i$ is the state index in basis $B^{(j)}_I$. This can be done in several ways. For example:

  - Thermal distribution with a given temperature. The distribution can be achieved over all states of the joined basis $B^{(j)}$. We can also perform a thermal distribution over a localized basis $B^{(j)}_I$, for example the injector region.
  - We can fix the population for any state in any basis $B^{(j)}_I$. This method is used to simulate special situations where fundamental modeling cannot be achieved. This method can be mixed with a thermal distribution.
  - The carrier concentration in each state is determined by a transport model. The details of its implementation are not required here.

- (j.3) The electron distribution is averaged with previous history for a few states.

  We sort all the states of the joined basis $B^{(j)}$ with increasing energies. This gives a basis $(E_k, \phi_k(z))$. We retain only the $K$ first states, the convergence will be made on them.

  We call $n^{(j)}_k$ and $E^{(j)}_k$ the electron concentration for the corresponding energy at iteration $j$. We determine the new damped populations and energies, $\overline{n}_k$ and $\overline{E}_k$ with:

  \[
  \overline{n}_k = N_j^{-1} \sum_{m=0}^{j} n^{(m)}_k e^{-(j-m)\alpha} \quad \overline{E}_k = N_j^{-1} \sum_{m=0}^{j} E^{(m)}_k e^{-(j-m)\alpha} \quad \text{with} \quad N_j = \sum_{m=0}^{j} e^{-m\alpha}
  \]

  for $k = 1..K$. The damping factor $\alpha$ needs to be chosen adequately for the structure under consideration. A standard value is $10^{-3}$.

- (j.4) From the averaged electron concentrations $\overline{n}_k$ corresponding to wavefunctions $\phi_k(z)$, a new self-consistent potential $\Phi^{(j+1)}$ is computed with Eq.(4.45). The net charge density $\rho^{(j+1)}$ is first computed. We force the periodicity in the fundamental period with Eq.(4.49) for each state.

- (j.5) If $j > 0$, we check if the computation converges, else we go to the next iteration first step ($(j+1)_1$).
We check convergence on energies only. We compute the variation of the damped energies $\delta E_k$:

$$
\Delta \bar{E} = \sqrt{\sum_{k=1}^{K} (\delta \bar{E}_k)^2} \quad \text{where} \quad \delta \bar{E}_k = \bar{E}_k^{(j)} - \bar{E}_k^{(j-1)}
$$

(4.54)

If the accuracy is reached we go to the final step, else we continue with step ((j+1).

- **(final)** We keep the last computed basis $B$, but we recompute the electron distribution with step (j.2): we do not use averaged populations or energies as a final result. They are only used for convergence.

Although the damping technique enables the convergence on a stable solution in most cases, the convergence is made slower for stable structures. Nevertheless as we give input populations to the self-algorithm that are not obtained by a simple model (like the thermal distribution) it is always better to use the damping technique in order to avoid a convergence-crash (reaching of the maximal number of iterations) in the middle of a chained computation.

**Improved self-consistent algorithm**

At the time of writing a much more efficient algorithm was proposed by J. Faist and we discuss it briefly here. The latter may have a large impact on the convergence of structures where the self-consistent contribution is large and depends on transport configurations.

The structure of the algorithm is basically the same as described above, however the damping procedure is not applied on the populations but on the self-consistent potential itself. The convergence remains however checked on the energies only.

Instead of a matrix of populations, we therefore have a matrix of potential energies. The latter are evaluated at each mesh points of the growth axis. Formally we have to compute as in Eq.(4.53) the damped values of the self-consistent potential $\Phi_k \equiv \Phi(z_k)$:

$$
\bar{\Phi}_k = N_j^{-1} \sum_{m=0}^{j} \Phi_k^{(m)} e^{-(j-m)\alpha}, \quad \text{for all values of } k.
$$

(4.55)

This new method is more reliable as, if the mesh is fixed, it is basically well-defined in any interesting cases. This was not the case previously with the populations as the number of states may varies if the self-consistent potential is large and/or unstable. Moreover the selection of the states themselves has to be performed very carefully.
Another reason of the success of this algorithm is that the populations may depend in a very complicated way of the total potential and they are subject to instabilities, which cause the damping technique to be less efficient and possibly to converge on a ill-defined solution (the resulting current-voltage curves are typically very noisy). Although the self-consistent potential is built from these populations, the converge on the potential itself is found more efficient. In a heuristic way, some portions of the self-consistent potential may converge more rapidly than others. However the blurring of some unstable portions will not yield a global unstable solution.

To give a telling example, we consider the structure shown in Fig.(4.7) that was designed in order to maximize the self-consistent potential. All the electron concentration was put in the ground-state. The standard algorithm based on population damping takes 62 iterations in order to converge with an accuracy of $1.5 \cdot 10^{-5}$ eV on the energies (taking into account all bound-states), while the improved algorithm takes only 3 iterations and converges with an accuracy of $2.8 \cdot 10^{-6}$ eV.

This improved self-consistent solver is a clear advantage when the transport populations have a strong impact on the potential shape of the structure. This is the case of the structure published in [99]. We give a few super-self-consistent computations based on this structure in (A).
Chapter 5

Non-radiative scattering between subbands

We consider non-radiative scattering mechanisms between conduction subbands. The goal of this section is to enable the calculation of the $\Gamma_{\text{inter}}$ and $\Gamma_{\text{intra}}$ energies broadening as described in [103] based on the work of T. Ando [2, 3]. The $\Gamma_{\text{inter}}$ basically gives the inter-subband scattering rate, while $\Gamma_{\text{intra}}$ gives the broadening of the linewidth due to dephasing between two concurrent intra-subband scattering. The intra-subband scattering rates (typically used in the computation of the electronic temperature (7.2)) are evaluated from the $\Gamma_{\text{inter}}$ terms by letting the same initial and final subbands. The pure phase contribution to the dephasing time (6.4.7) will be evaluated using $\Gamma_{\text{intra}}$.

Important assumption

For all interactions, we assume that the envelop functions associated with the subbands are computed in the effective two-band model presented in (3.2). The wavefunctions are therefore not normalized to unity since they are only conduction-band components.

5.1 Definition of $\Gamma_{\text{inter}}$ and $\Gamma_{\text{intra}}$

In the next paragraph we follow the paper of T. Unuma [103] for the main definitions. We consider the scattering part of the hamiltonian $H_{\text{scatt}}$. We treat each scattering potential independently in the Fermi’s Golden rule approximation. We therefore neglect scattering terms that may arise from interferences between the various scattering mechanisms.

The state of subband $\mu$ with in-plane wave-vector $\mathbf{k}$ is denoted by $\ket{\mu \mathbf{k}}$. The subbands
are computed in the envelop function approximation. The in-plane energy-dispersion is obtained to the first order with an effective mass $m_\mu$.

Figure 5.1: The $\Gamma_{\text{inter}}$ and $\Gamma_{\text{intra}}$ energy broadening are illustrated. The initial wave-vector is $k$, the final wave-vector is $k'$, the exchanged momentum is $q$. The inelastic scattering is considered with the emission of one quantum $\hbar\omega_0$. (a) Inter-subband elastic scattering. (b) $\Gamma_{\text{intra}}$ for elastic scattering. (c) Inter-subband inelastic scattering (emission). (d) $\Gamma_{\text{intra}}$ for inelastic scattering (emission).
The intra-subband energy broadening and the inter-subband scattering are illustrated in Fig.(5.1).

5.1.1 Definition of $\Gamma_{\text{inter}}$

The energy broadening due to inter-subband scattering from state $|\nu k\rangle$ to all possible states $|\mu k'\rangle$ for a particular interaction $H_{\text{scatt}}$ reads:

$$\Gamma_{\text{inter}}^{\nu \rightarrow \mu} = 2\pi \sum_{k'} \langle | \langle \mu k' | H_{\text{scatt}} | \nu k \rangle |^2 \rangle \delta\left(\epsilon_{\nu}(k) - \epsilon_{\mu}(k') + E_{\nu\mu}\right)$$

(5.1)

where the subband energy-dispersion is given by:

$$\epsilon_{\nu}(k) = \epsilon_{\nu}(0) + \epsilon_{\nu}(k) \quad \text{with} \quad \epsilon_{\nu}(k) = \frac{\hbar^2 k^2}{2m_{\nu}}$$

(5.2)

$\epsilon_{\nu}(k)$ being the kinetic energy. The subband detuning energy is $E_{\nu\mu} \equiv \epsilon_{\nu}(0) - \epsilon_{\mu}(0)$. The notation $\langle ... \rangle$ denotes the average on the distribution of scatterers (if applies).

In Eq.(5.1) we have not multiplied the transition probability by the occupancy function of the final subband. The evaluation of averaged scattering rates is discussed later in (5.2.1).

The sum over final states $|\mu k\rangle$ is performed here to extract the energy conservation from the Dirac function. We could have written: $\Gamma_{\text{inter}}^{\nu \rightarrow \mu} = 2\pi \langle | \langle \mu k_f | H_{\text{scatt}} | \nu k \rangle |^2 \rangle$, where $k_f$ is solution of the implicit equation: $\epsilon_{\nu}(k) = \epsilon_{\mu}(k_f)$.

In Eq.(5.1) we assumed that the scattering was elastic and therefore yields conservation of the energy $\delta\left(\epsilon_{\nu}(k) - \epsilon_{\mu}(k')\right)$. For inelastic scattering (LO-phonon for example) we assume that at the first order a quantum of $\hbar \omega_0$ is emitted or absorbed. The previous delta function is replaced by $\delta\left(\epsilon_{\nu}(k) - \epsilon_{\mu}(k') \pm \hbar \omega_0\right)$, where the (+) is for absorption and the (−) for emission.

5.1.2 Definition of $\Gamma_{\text{intra}}$

If we now consider the energy broadening due to dephasing between intra-subband scattering in two subbands $\mu$ and $\nu$, we can write in the parabolic approximation:

$$\Gamma_{\text{intra}}^{(\mu,\nu)} = 2\pi \sum_{k'} \langle | \langle \mu k' | H_{\text{scatt}} | \mu k \rangle - \langle \nu k' | H_{\text{scatt}} | \nu k \rangle |^2 \rangle \delta\left(\epsilon(k) - \epsilon(k')\right).$$

(5.3)

This energy broadening is defined for a pair of states as outlined by the notation $(\mu, \nu)$. The Dirac function imposes the energy conservation in each subband. The sum is performed on final states to extract the energy conservation from the Dirac function.
In the above expression we have assumed that $m_\mu = m_\nu \equiv m_w$ where $m_w$ is the effective mass of each subband and therefore the kinetic energy $\epsilon(k)$ has the same expression for each subband. This expression is of course difficult to apply to III-V alloys where the non-parabolicity is expected to be strong. With non-parabolicity $\Gamma_{\text{intra}}$ can be expressed as:

$$
\left| \langle \mu k' | H_{\text{scatt}} | \mu k \rangle \right|^2 \delta(\epsilon_\mu(k) - \epsilon_\mu(k')) + \left| \langle \nu k' | H_{\text{scatt}} | \nu k \rangle \right|^2 \delta(\epsilon_\nu(k) - \epsilon_\nu(k')) + 2\Re \left[ \langle \mu k' | H_{\text{scatt}} | \mu k \rangle \langle \nu k' | H_{\text{scatt}} | \nu k \rangle \right] \frac{1}{2} \left( \delta(\epsilon_\mu(k) - \epsilon_\mu(k')) + \delta(\epsilon_\nu(k) - \epsilon_\nu(k')) \right) 
$$

which require a more careful evaluation. However in the calculations performed without non-parabolicity, the individual contributions from each subband are clearly identifiable. It enables to set the corresponding Dirac functions before the right terms.

We have considered an elastic process. For an inelastic process, the Dirac function should be modified in $\delta(\epsilon_\mu(k) - \epsilon_\mu(k') \pm \hbar \omega_0)$ and $\delta(\epsilon_\nu(k) - \epsilon_\nu(k') \pm \hbar \omega_0)$, where (+) is for absorption and (−) for emission.

It is worth to remark that $\Gamma_{\text{inter}}^{\nu \rightarrow \mu}(k)$ and $\Gamma_{\text{intra}}^{(\mu, \nu)}(k)$ depend on the initial wave-vector $k$ from the initial state $|\nu k\rangle$. Even with non-parabolicity it is possible to parametrize $\Gamma_{\text{inter}}^{\nu \rightarrow \mu}$ with the in-plane kinetic energy of the initial subband $\epsilon_\nu(k)$. But for $\Gamma_{\text{intra}}^{(\mu, \nu)}(k)$ this parametrization is meaningless. We therefore prefer to keep the dependence in the initial wave-vector $k$.

### 5.2 Scattering rates, dephasing times and optical linewidths

The transport model developed in (7) requires the knowledge of: the inter- and intra-scattering rates, the dephasing times and later (8) the optical linewidth for each possible transitions. All these quantities are related to $\Gamma_{\text{inter}}$ and $\Gamma_{\text{intra}}$ energies.

We assume that we have computed the contributions of all relevant scattering mechanisms: the latter will be detailed in (5.3-5.4). Since we are using a first order approximation, the total energy broadening is the sum of each individual broadening. We can write it formally as:

$$
\Gamma_{\text{inter}} = \sum_X \Gamma_X^{\text{inter}} \quad \text{and} \quad \Gamma_{\text{intra}} = \sum_X \Gamma_X^{\text{intra}}
$$

where $X$ corresponds to a particular scattering mechanism.
5.2. Scattering rates, dephasing times and optical linewidths

In the following section we define relevant quantities for the transport model derived in (7). In particular as the later is an effective 0-dimensional model, we have to simplify the \( k \)-dependence of the broadening energies.

5.2.1 Calculation of the inter-subband scattering rates

The inter-subband scattering rate \( W_{\nu \rightarrow \mu}(k) \) from state \(|\nu k\rangle\) to subband \( \mu \) is obtained from the \( \Gamma_{\text{inter}}^{\nu \rightarrow \mu} \) term:

\[
W_{\nu \rightarrow \mu}(k) \equiv \frac{1}{\hbar} \Gamma_{\text{inter}}^{\nu \rightarrow \mu}(k).
\]  

(5.6)

The \( k \)-dependency has to be removed by integration over an electron distribution in the initial subband \( \nu \) or by evaluating the scattering rate at a fixed wave-vector \( k_0 \). The latter being usually either zero (scattering rate evaluated at the subband-edge) with \( k_0 = 0 \) or fixed by a thermal energy with \( k_0 = \hbar^{-1} \sqrt{2m_\nu k_b T_\nu} \), where \( k_b \) is the Boltzmann constant and \( T_\nu \) the electronic temperature of the initial subband. We implicitly assumed here thermal distributions in the subbands. These simple approximations were implemented in order to test the model numerically, however we have finally implemented the averaging over thermal distributions in subbands as explained below and in (6.1.2).

We assume that the electrons are distributed in each subband \( \nu \) with a distribution \( f_\nu(k) \). The latter can be a Fermi-Dirac distribution as in (6.1.2). The total population in subband \( \nu \) is given by:

\[
N_\nu = \sum_k f_\nu(k). 
\]  

(5.7)

The averaged scattering rate accounting for the occupancy of the final subband \( \mu \) is given by:

\[
\bar{W}_{\nu \rightarrow \mu} = \frac{1}{N_\nu} \sum_k \sum_{k'} W_{\nu \rightarrow \mu}(k)f_\nu(k)(1 - f_\mu(k'))
\]  

(5.8)

This scattering rate is used in 0-dimensional models. The standard distribution is the Fermi-Dirac function.

The continuum-limit is usually taken and the sums are replaced by integrals: \( \sum_k \rightarrow \int d^2k \) \( D_\nu \), where \( D_\nu \) is the density of states for subband \( \nu \).
Chapter 5. Non-radiative scattering between subbands

5.2.2 Calculation of the dephasing times

As examined in (6.4.7), the dephasing time between a pair of subbands \((\nu, \mu)\) can be expressed in energy units \((\gamma = \hbar/\tau_\perp)\) by:

\[
\gamma(\nu, \mu) = \gamma^{(0)}(\nu, \mu) + \frac{1}{2} \left( \Gamma^\mu_{\text{inter}} + \Gamma^\nu_{\text{inter}} \right). \tag{5.9}
\]

The term \(\gamma^{(0)}(\nu, \mu)\) is a pure phase contribution as it does not arise from inter-subband scattering.

This term appears naturally in the density matrix formalism as shown in Eq.(29) in [55].

In this model we have:

\[
\gamma(\nu, \mu) = \frac{1}{2} \Gamma^{(\nu, \mu)}_{\text{intra}} + \frac{1}{2} \left( \Gamma^\mu_{\text{inter}} + \Gamma^\nu_{\text{inter}} \right) \tag{5.10}
\]

The expression coincides with the dephasing time in the Ando model [2]. However the computation in this work were performed with a different expression. For the current density, the difference between the results is about 5-10 %.

The dephasing energy \(\gamma(\nu, \mu)\) depends on \(k\). In order to remove this dependency, we evaluate Eq.(5.10) at the thermal energy by assuming thermal intrasubband-distributions in the subbands. We have:

\[
k_{\text{th}} = \hbar^{-1} \sqrt{2m_\nu k_b T_e^{(\nu)}} \]

that we substitute in Eq.(5.4).

The second contribution in Eq.(5.9) arises from the total scattering rate of state \(\mu\) and state \(\nu\). The latter is a sum over all possible final subbands:

\[
\Gamma^\nu_{\text{inter}} = \sum_{\mu \neq \nu} \Gamma^{\nu \rightarrow \mu}_{\text{inter}}. \tag{5.12}
\]

The pure-phase contribution is overestimated. We have considered this expression by following the literature where the phase relaxation time in nuclear magnetic resonance (NMR) models is given by:

\[
1/T_1 = 1/T_2 + 1/(2T_1),
\]

where \(T_1\) is the spin-lattice relaxation (which is mapped on inter-subband transitions) and \(T_2\) is the spin-spin relaxation (we abusively mapped on \(\Gamma^{(\nu, \mu)}_{\text{intra}}\)).

At the time of writing we have used expression in Eq.(5.11). However we have corrected the evaluation of the dephasing time in our numerical simulator by implementing Eq.(5.10).

The variations of the computed maximal current densities is about 5-10% for the structures investigated in this work. This is discussed in (B) in more details.
5.2.3 Calculation of the optical linewidth

In the Ando model [103], the optical line width for a transition between subband $\nu$ and subband $\mu$ is given by the sum:

$$\Gamma_{\text{opt}}^{\nu \rightarrow \mu}(k) = \frac{1}{2} \left( \Gamma_{\text{intra}}^{\nu,\mu}(k) + \Gamma_{\text{inter}}^{\nu \rightarrow \mu}(k) + \Gamma_{\text{inter}}^{\mu \rightarrow \nu}(k) \right).$$

(5.13)

For a lorentzian (symmetric) optical line $\Gamma/(\delta^2 + \Gamma^2)$ as considered in (8.2), the FWHM (Full-Width-At-Half-Maximum) is given by $2\Gamma_{\text{opt}}$.

The optical linewidth coincides with the dephasing energy given in Eq.(5.10). It is normal as the gain and the current are driven by coherences. The calculations given in (6.4.7) for an arbitrary coherence $\rho_{ab}$ therefore hold both for the gain and the current.

5.3 Elastic scattering mechanisms

We examine elastic scattering mechanisms relevant for mid-infrared quantum cascade structures. The most important is the interface roughness scattering. We also discuss alloy disorder scattering (the latter can be viewed as a limit of interface roughness scattering) and scattering on ionized impurities (dopants).

5.3.1 Interface roughness

The interface roughness is mainly caused by growth imperfections at the junction of two semiconductor alloys. During the growth, the deposition of one alloy over the surface formed by the other, causes in about one monolayer a fluctuation in the composition of the interface.

For now we consider only one interface [3, 33]. We assume the height of the roughness to be given by a function $\Delta(r)$ where $r = (x, y)$ are the in-plane coordinates. We model the $\Delta(r)$ function by its auto-correlation function. We assume that the latter is gaussian:

$$\langle \Delta(r) \Delta(r') \rangle = \Delta^2 \exp \left( -\frac{|r - r'|^2}{\Lambda^2} \right)$$

(5.14)

The parameters are the mean height $\Delta$ of the roughness and the in-plane correlation length $\Lambda$. The latter means that the distance between similar composition fluctuations follows a gaussian law, the mean distance being given by $\Lambda$. 

The conduction-band edge between the two semiconductor alloys have an offset: $+V_0$ when the interface is made between the well-acting material (left) and the barrier acting material (right). The offset being $-V_0$ between barrier- and well-acting materials.

The fluctuation of the interface position therefore implies a fluctuation of the energy. The matrix element between two subbands is given by:

$$\langle \mu k | H_{IFR} | \nu k \rangle = \int d^2 r F_{\mu \nu} \Delta(r) \exp(\imath \mathbf{q} \cdot \mathbf{r}) \quad (5.15)$$

where $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ is the exchanged momentum. The strength of the interaction is given by:

$$F_{(\mu,\nu)} = V_0 \chi_\nu(z_0) \chi_\mu(z_0) \quad (5.16)$$

where $V_0$ is the potential drop between well and barrier materials, $\chi_\mu(z)$ is the $z$-dependent part of the envelop function of subband $\mu$ and $z_0$ is the position of the interface. Before we generalize this results to several correlated interfaces, we examine only one interface. It is worth to remark that $F$ is defined for a pair of states $(\mu, \nu)$ as it is symmetric under the exchange of subband indices.

In his study of the optical linewidth [103], T. Unuma have considered a potential well of width $L$ centered at $z = 0$. The potential difference involved in the definition of $F_{(\mu,\nu)}$ is positive for a negative potential drop computed from left to right. This convention arises directly from the computation of the variation of the energy [3] and we will not report the complete derivation here. The strength of the interface roughness scattering $F_{(\mu,\nu)}$ can be computed as the variation of the subband-edge energies with respect of the well width $L$. The interface roughness basically causes fluctuations of the well width:

$$F_{(\mu,\nu)} = \sqrt{\left| \frac{\partial \varepsilon_\mu(0)}{\partial L} \right| \left| \frac{\partial \varepsilon_\nu(0)}{\partial L} \right|} \quad (5.17)$$

This formulation is not very useful for computing the interface roughness scattering in a system with several interfaces like quantum cascade structures. However it gives some insight in the scaling of the interface roughness scattering with well-widths. It is interesting to estimated the strength ratio of this interaction between mid- and far-infrared structures. If we consider states in an infinite well, the eigen-energies are given by: $E_n = \frac{n^2 \hbar^2 \pi^2}{2mL^2}$. Between two states $F$ scales as:

$$F \propto L^{-3} \quad (5.18)$$

The energy broadening is proportional to $F^2$. The typical active-well width for far-infrared structures is $L_{\text{FIR}} \approx 28$ nm. It is $L_{\text{MIR}} \approx 7.2$ nm for mid-infrared structures. The ratio of $F^2$ values is given by:

$$\left( \frac{F_{\text{FIR}}}{F_{\text{MIR}}} \right)^2 = \left( \frac{L_{\text{MIR}}^3}{L_{\text{FIR}}^3} \right)^2 \approx 2.8 \cdot 10^{-4}. \quad (5.19)$$
5.3. Elastic scattering mechanisms

If interface roughness basically dominate elastic scattering for mid-infrared quantum cascade structures, it is not the case in the far-infrared. In the far-infrared typical linewidth are about 1 meV, while in the mid-infrared it is 15 meV. The ratio is therefore $10^{-1}$ and not $10^{-4}$. The linewidth in far-infrared structures cannot therefore be explained by interface roughness scattering considered as an homogeneous (lifetime) broadening.

**Interface roughness as a source of inhomogeneous broadening** Interface roughness can also be considered as a source of inhomogeneous broadening. We consider the islets of diameter $\Lambda$ formed by interface roughness in the in-plane dimensions. In these islets the well-widths vary and therefore the energies of the quantum states in the (growth) $z$-direction. The typical energy variation is given by $\delta E = \delta L/L^3$. If we assume that these fluctuations are uncorrelated (inhomogeneous), the effective linewidth will be give by an incoherent superposition of homogeneously broadened optical transitions. The distribution will be typically gaussian where the mean energy is given by the computed bare energy with perfect interfaces and the standard-deviation is given by $\delta E$.

The homogeneous broadening due to interface roughness scales with $F^2 \propto L^{-6}$. However the inhomogeneous contribution scales with $L^{-3}$. There is a cross-over. As the linewidth given by homogeneous broadening will decreases as $L^{-6}$, the interface roughness will progressively appear as an inhomogeneous broadening source.

**Inter-subband energy broadening**

In order to fix the definition of the $F_{(\mu,\nu)}$ elements we adopt the convention [101]:

$$F_{(\mu,\nu)} = -\delta V(z_i) \chi_{\mu}(z_i) \chi_{\nu}(z_i) \quad \text{where} \quad \delta V(z_i) = V(z_i^+) - V(z_i^-). \quad (5.20)$$

where $V(z)$ is the (conduction) band-edge profile and $z_i^+$ means values of $z_i$ approached with $z > z_i$ and $z_i^-$, values of $z_i$ approached with $z < z_i$.

The value of $\Gamma_{\nu \rightarrow \mu}^{\text{int}}$ is obtained by substituting Eq.(5.15) in Eq.(5.1). The conservation of energy imposed by the Dirac function determines the $k'$ values:

$$k'^2 = \frac{m_{\mu}}{m_{\nu}} k^2 + \frac{2m_{\mu} E_{\nu \mu}}{\hbar^2} \quad (5.21)$$

The squared norm value of the exchanged momentum $q^2$ is therefore given by $q^2 = k^2 + k'^2 - 2kk' \cos \theta$ as:

$$q^2 = \left( 1 + \frac{m_{\mu}}{m_{\nu}} \right) k^2 + \frac{2m_{\mu} E_{\nu \mu}}{\hbar^2} - 2k \sqrt{\frac{m_{\mu}}{m_{\nu}}} k^2 + \frac{2m_{\mu} E_{\nu \mu}}{\hbar^2} \cos \theta \quad (5.22)$$
When $m_\mu = m_\nu$, the standard formula is recovered. The transition probability can be computed by changing to polar coordinates in the plane of the heterostructure and by identifying the Fourier transform of the correlation function. The conservation of the energy requires to express the Dirac function as:

$$
\delta \left[ \frac{\hbar^2}{2m_\nu} \left( k^2 - \frac{m_\nu}{m_\mu} k'^2 + \frac{2m_\nu E_{\nu\mu}}{\hbar^2} \right) \right] = \frac{2m_\nu}{\hbar^2} \delta \left[ k^2 - \frac{m_\nu}{m_\mu} k'^2 + \frac{2m_\nu E_{\nu\mu}}{\hbar^2} \right]
$$

(5.23)

After some careful calculations one can obtain:

$$
\Gamma_{\text{inter}}^{\nu\rightarrow\mu} = \frac{m_\nu \Delta^2 \Lambda^2}{\hbar^2} F^2_{(\mu,\nu)} \int_0^\pi d\theta \exp \left( -\frac{q^2 \Lambda^2}{4} \right),
$$

(5.24)

where the exchanged momentum is given in Eq.(5.22).

The form factor given by the integral depends strongly on the energy detuning between subbands $E_{\nu\mu}$. If we consider a simple case by letting $k = 0$ (the initial state is the edge of the initial subband), we have: $q^2 = 2m_\nu E_{\nu\mu}/\hbar^2$. A final state exists only if $q^2 \geq 0$. The integral in the form factor $f(E)$ is trivial and is given for $E \equiv E_{\nu\mu} \geq 0$ by:

$$
f(E) = \pi \exp \left( -\frac{2m_\mu \Lambda^2}{4\hbar^2} E \right)
$$

(5.25)

The energy broadening therefore decays exponentially with the energy.

**Intra-subband energy broadening**

We now consider the contribution of interface roughness scattering to intra-subband broadening. The latter is obtained by similar calculations used for the inter-subband term. We have to take care of each term in Eq.(5.4) when non-parabolicity is included. The strength of the interaction is given by $F_{(\mu,\mu)} = V_0 \chi_\mu(z_0)$ and $F_{(\nu,\nu)} = V_0 \chi_\nu(z_0)$. The exchanged momentum imposed by $\delta(\epsilon_\mu(k) - \epsilon_\mu(k'))$ and $\delta(\epsilon_\nu(k) - \epsilon_\nu(k'))$ is the same and it reads:

$$
q^2 = 2k^2 (1 - \cos \theta).
$$

(5.26)

We have:

$$
\Gamma_{\text{intra}}^{(\mu,\nu)} = \frac{\Delta^2 \Lambda^2}{\hbar^2} \left[ m_\mu F^2_{(\mu,\mu)} + m_\nu F^2_{(\nu,\nu)} - (m_\mu + m_\nu) F_{(\mu,\nu)} F_{(\nu,\nu)} \right] \int_0^\pi d\theta \exp \left( -\frac{q^2 \Lambda^2}{4} \right).
$$

(5.27)

Each term is weighted by the mass of the corresponding subband. The form factor does not depend on the energy-separation between the subbands as expected.
5.3. Elastic scattering mechanisms

Calculation of the interface roughness coupling constants $F^2$ for inter-subband energy broadening in a system with several interfaces

We consider the calculation of interface roughness scattering for an arbitrary number of interfaces. We assume that a potential $V(z)$ has $M$ interfaces located at positions $z_i$. For each of these interfaces the coupling strength $F_{(\mu,\nu)}$, for two subbands $\mu$ and $\nu$ is given by (we omit the label \textit{inter}):

$$F_{(\mu,\nu)}^{(i)} = -\delta V(z_i) \chi_\mu(z_i) \chi_\nu(z_i) \text{ where } \delta V(z_i) = V(z_i^+) - V(z_i^-). \quad (5.28)$$

for $i = 1, \ldots, M$. We may assume that these interfaces are correlated in the vertical (growth) direction [101] by replacing the squared coupling strength $F^2$ in Eq.(5.24) by:

$$\mathcal{R}_{(\mu,\nu)}^2 = \sum_{i,j} c_{ij} F_{(\mu,\nu)}^{(i)} F_{(\mu,\nu)}^{(j)}. \quad (5.29)$$

For uncorrelated interfaces we have $c_{ij} = \delta_{ij}$ and $\mathcal{R}_{(\mu,\nu)}^2$ is a sum of squared coupling strength for each interfaces. However we may assume an exponential correlation in the growth direction $z$:

$$c_{ij} = \exp\left(-\frac{|z_i - z_j|}{\kappa}\right). \quad (5.30)$$

The diagonal terms is still unity $c_{ii} = 1$, but there exist a correlation between the interface that decays over a characteristic length of $\kappa$: \textit{the vertical correlation length}. If $\kappa \to 0$, we recover the uncorrelated case, but if $\kappa \to \infty$, the interfaces are perfectly correlated. In general this not yields a zero coupling strength, but the magnitude of $\mathcal{R}_{(\mu,\nu)}^2$ can be largely reduced.

This vertical correlation model is not relevant in most cases as these correlations are not observed for interfaces separated by a distance larger than $\approx 1.5$ nm [101]. However this model allows to remove artificial interface scattering that arises from vey thin barriers (spikes) that have a thickness of 0.2 nm (below one mono-layer). If we assume uncorrelated interfaces, each interface of the spike contributes independently to scattering. This approximation is very bad as for a width under one monolayer we can assume that both interfaces are perfectly correlated. We assume that the values of the wavefunctions are the same on both side of the spike located at $z_0 \pm \delta z$, where $2\delta z$ is the thickness of the spike.

We have $F_1 = -V_0 \sqrt{f}$ and $F_2 = V_0 \sqrt{f}$, where $f = \chi_\mu^2(z_0)\chi_\nu^2(z_0) \geq 0$ is the wavefunctions product. The uncorrelated interfaces model gives a coupling strength of:

$$\mathcal{R}_{(\mu,\nu)}^2 = F_1^2 + F_2^2 \approx 2V_0^2 \chi_\mu^2(z_0)\chi_\nu^2(z_0). \quad (5.31)$$
where $V_0$ is the discontinuity between well and barriers. If we now apply the vertical correlation model, we have:

$$\mathcal{F}^2_{(\mu,\nu)} = F_1^2 + F_2^2 + 2cF_1F_2 \approx (1 - c)2V_0^2f$$

(5.32)

If the correlation is perfect: $c = 1$ and $\mathcal{F}^2_{(\mu,\nu)} = 0$. The effect of the spike has been removed. However this model is too strong as the spike contributes to scattering as one-effective interface. For a finite $\kappa$ we have $c < 1$. And in general the values of the wavefunctions at each pseudo-interfaces $z_0 \pm \delta z$ are different.

We choose $\kappa = 1.5$ nm for typical mid- (and far-) infrared structures. The other interfaces are typically uncorrelated but the impact of the spike is decreased. We assume to model correctly the one-effective interfaces given by the spikes. The intra-subband energy broadening will also benefits from a vertical correlation model (5.3.1). However $\Gamma_{\text{intra}}$ has a strong impact on the linewidth, in particular of the luminescence. Measurements of linewidths in single-quantum-well regions in the mid-infrared with spike in the active region are supporting the vertical correlation model discussed above.

**Calculation of the interface roughness coupling constants $F^2$ for intra-subband energy broadening in a system with several interfaces**

As in the case of inter-subband broadening, we are considering the $F_{(\mu,\nu)}$ defined for $M$ interfaces:

$$F^{(i)}_{(\mu,\nu)} = -\delta V(z_i) \chi_\mu(z_i)\chi_\nu(z_i) \quad \text{where} \quad \delta V(z_i) = V(z_i^+) - V(z_i^-).$$

(5.33)

for $i = 1, \ldots, M$. In the case of intra-subband interface roughness, the coupling constants have typically the form $[F_{(\mu,\mu)} - F_{(\nu,\nu)}]^2$ for a pair of subbands $\mu$ and $\nu$. We will treat the case of non-parabolicity at the end of this section. We can write the squared coupling strength as (we omit the label $\text{intra}$):

$$\mathcal{F}^2_{(\mu,\nu)} = \sum_{i,j}^M c_{ij} \left( F^{(i)}_{(\mu,\mu)} - F^{(i)}_{(\nu,\nu)} \right) \left( F^{(j)}_{(\mu,\mu)} - F^{(j)}_{(\nu,\nu)} \right)$$

(5.34)

As before, we have introduced a correlation matrix $c_{ij}$. The uncorrelated case, yielding a sum of squared coupling strengths, is obtained with $c_{ij} = \delta_{ij}$. We consider the previously defined exponential vertical correlation with a characteristic length $\kappa$. The discussion about spikes can be developed here with similar results as the one obtained in the inter-subband case.
A more complicated question is the implementation of non-parabolicity. As shown in Eq. (5.27), each part of the squared coupling strength is weighted with the mass of the corresponding subband, as the non-parabolicity is very strong in mid-infrared quantum cascade structure, we have to implement non-parabolicity in the vertical correlation model. We expand the product under the sum in Eq. (5.34):

\[
(F^{(i)}_{(\mu,\mu)} - F^{(i)}_{(\nu,\nu)})(F^{(j)}_{(\mu,\mu)} - F^{(j)}_{(\nu,\nu)}) = F^{(i)}_{(\mu,\mu)} F^{(j)}_{(\mu,\mu)} + F^{(i)}_{(\nu,\nu)} F^{(j)}_{(\nu,\nu)} - F^{(i)}_{(\mu,\mu)} F^{(j)}_{(\nu,\nu)} - F^{(i)}_{(\nu,\nu)} F^{(j)}_{(\mu,\mu)}.
\] (5.35)

We attribute to these terms the corresponding masses, following Eq. (5.4). The squared coupling strength reads:

\[
F^2_{(\mu,\nu)} = \frac{m_{\mu}}{m_{\mu} + m_{\nu}} \sum_{i,j} c_{ij} F^{(i)}_{(\mu,\mu)} F^{(j)}_{(\mu,\mu)} + \frac{m_{\nu}}{m_{\mu} + m_{\nu}} \sum_{i,j} c_{ij} F^{(i)}_{(\nu,\nu)} F^{(j)}_{(\nu,\nu)} - \sum_{i,j} c_{ij} (F^{(i)}_{(\mu,\mu)} F^{(j)}_{(\nu,\nu)} + F^{(i)}_{(\nu,\nu)} F^{(j)}_{(\mu,\mu)}).
\] (5.36)

This expression holds, provided that we modify the expression of the energy broadening \(\Gamma^{(\mu,\nu)}_{\text{intra}}\) as:

\[
\Gamma^{(\mu,\nu)}_{\text{intra}} = \frac{\Delta^2 \Lambda^2}{\hbar^2} (m_{\mu} + m_{\nu}) F^2_{(\mu,\nu)} \int_0^\pi d\theta \exp \left(-\frac{q^2 \Lambda^2}{4}\right). \] (5.37)

**Vertical correlation: a numerical example**

We give a numerical example by calculating the intra-subband broadening in a mid-infrared single-quantum-well active region. The active region consists of two wells of 3.6 nm separated by a spike of 2 Å. The material system is Ga_{0.47}In_{0.53}As /Al_{0.48}In_{0.52}As lattice matched on InP. The band-structure is shown at alignment (when the whole period is considered) electric field (\(F = 48 \text{ kV/cm}^2\)) in Fig. (5.2a).

The intra- and inter-subband broadening are computed for typical parameters \(\Delta = 1.2 \text{ Å}\) and \(\Lambda = 9.0 \text{ nm}\). The vertical correlation length \(\kappa\) was varied from 0 Å to 100 Å in order to illustrate its impact on the energies. The results are shown in Fig. (5.2c). For low \(\kappa\) values between 0 and 10 Å , the intra-subband broadening is largely over-estimated as show in the zoomed panel at the left of the figure. For \(\kappa \leq 1 \text{ Å}\) the linewidth is 43 meV, while it is about 7.2 meV for \(\kappa \leq 15 \text{ Å}\); the value we are using for the vertical correlation length. This value of 7.2 meV is of the same order of magnitude (and is numerically very close) of the measured linewidth. We have also reported the intra-subband broadening in blue for a single-quantum-well region without spikes: its value in the plateau region is \(\approx 8 \text{ meV}\) which is very close to the previous value of 7.2 meV. The modulation of the intra-subband energy broadening is therefore clearly attributed to the pseudo-interfaces of the spike.
Figure 5.2: The effect of vertical correlation is examined in a single-quantum-well active region (N258). The details are given in the text. The layer sequence of the structure is given by: 4.8/3.6/0.2/3.6/3.5/5.1/1.1/5.0/1.2/4.5/1.3/3.5/1.5/3.4/1.6/3.3/1.8/3.2/2.1/3.0/2.5/3.0/2.9/2.9. Barriers are in bold and wells in roman.
The lifetime of the first excited state, the lifetime is under-estimated for low \( \kappa \) values. The curves in Fig.(5.2c) show a plateau around \( \kappa \leq 15 \text{ Å} \) : this is the physical region of the system. The intra-subband broadening then shows a local maximum near the well-width of 3.6 nm as the spike contribution is partially canceled. For the inter-subband scattering time a maximum is reached near 50 Å. The interpretation is less simple as the sign of the wavefunctions is important in the cancelation.

We consider the injector region of the same structure as shown in Fig.(5.2b). There is no spike in the latter and it is therefore a good test-ground for the vertical correlation model. The latter should not affect too much the energy broadening of these states. We have chosen state #5 and #4 (counted from the ground-state of the injector) because they span a large number of interfaces. The results are shown in Fig.(5.2d). It is clear that \( \kappa \leq 15 \text{ Å} \) is the limit value as the intra-subband broadening and the lifetime of state #5 begin to change rapidly for higher values of \( \kappa \).

The vertical correlation model in quantum cascade structures seems to be efficient for reasonable numerical values of the correlation parameter.

The computed values are in a good agreement with measurements (luminescence linewidth) and published results [103]. However the vertical correlation with spikes has not been investigated (to our best knowledge) and it is partially the reason of these calculations. It also shows that the interface roughness scattering is correctly implemented. It is worth to mention that the accuracy of the interface-roughness modeling is crucial for mid-infrared quantum cascade structures [57].

### 5.3.2 Alloy disorder

In this section we consider the inter-subband scattering and the intra-subband energy broadening due to alloy disorder. We consider only alloys of type \( \text{A}_x\text{B}_{1-x}\text{C} \) like \( \text{Ga}_{0.47}\text{In}_{0.53}\text{As} /\text{Al}_{0.48}\text{In}_{0.52}\text{As} \) lattice matched on \( \text{InP} \) (for mid-infrared) or \( \text{Al}_x\text{Ga}_{1-x}\text{As} \) (for far-infrared \( x = 0.15 \)). However the model can be applied and yields accurate results for stained-compensated materials as shown in (5.5).

Alloy disorder can be considered as a limit case of the interface roughness scattering, although it has been treated independently [7,103].

**Calculation of the \( \Gamma_{\text{inter}} \) contribution**

We could have followed [7] by computing directly the scattering rates due to alloy disorder. The origin of the latter is the *random contamination* of A atoms in BC alloy, or respectively
of B atoms in AC alloy. The model can even be extended to higher order alloys, but here
we consider the simplest case. We obtain alloy disorder scattering as a limit case of
interface roughness scattering.

The presence of a fraction \( f_A \) of atoms A in the bulk alloy BC, causes random scattering of
Bloch waves in the planar dimensions of the heterostructure. In a first order approximation,
this yields incoherent scattering between subbands. If we consider the expression for \( \Gamma_{\mu \rightarrow \nu}^{\text{IFR}} \)
from subband \( \mu \) to subband \( \nu \). We have dropped the \textit{inter} label for clarity. The Eq.(5.24)
reads:

\[
\Gamma_{\nu \rightarrow \mu}^{\text{inter}} = \frac{m_\nu \Delta^2 \Lambda^2}{\hbar^2} F_{(\mu, \nu)}^2 \int_0^\pi d\theta \exp \left( -\frac{q^2 \Lambda^2}{4} \right) \text{ with } F_{(\mu, \nu)}^2 = \delta V(z_i) \chi_\mu^2(z_i) \chi_\nu^2(z_i),
\]

where the potential drop is \( \delta V(z_i) = V(z_i^+) - V(z_i^-) \equiv \delta E_c \), with \( \delta E_c \) being the conduction
band discontinuity between the alloy AC and BC. The exchanged momentum is given by:

\[
q^2 = \left( 1 + \frac{m_\mu}{m_\nu} \right) k^2 + \frac{2m_\mu E_{\nu \mu}}{\hbar^2} - 2k \sqrt{\frac{m_\mu k^2 + \frac{2m_\mu E_{\nu \mu}}{\hbar^2}}{\cos \theta}}
\]

(5.39)

Now we assume that the interface roughness has an atomic size by letting \( \Delta = a \), where
\( a \) is the lattice constant of the material system: we assume that AC and BC are lattice
matched. We also assume that the roughness has no in-plane correlation by letting \( \Lambda = a \).
Of course we are neglecting any vertical correlations: \( \kappa = a \) and therefore adjacent mono-
layers are uncorrelated.

The form factor is given by the integral in Eq.(5.38). If we let \( \Lambda = a \), then the product
\( q^2 a^2 \ll 1 \) since we are comparing crystal momentum in the heterostructure with
atomic distances. For any elastic scattering between subbands this condition holds:
\( \exp \left( -q^2 \Lambda^2/4 \right) \approx 1 \) and therefore, the form factor is equal to \( \pi \).

The Eq.(5.38) therefore reads:

\[
\Gamma_{\nu \rightarrow \mu}^{\text{inter}} = \frac{m_\nu a^3 (\delta E_c)^2}{\hbar^2} f_A f_B \pi a \langle \chi_\mu^2 \chi_\nu^2 \rangle_{\text{ML}}
\]

(5.40)

where we have multiplied the result by the alloy fraction \( f_A = x \) and \( f_B = 1 - x \). We have
introduced an average \( \langle \ldots \rangle \) in order to underline that the value of the product \( \chi_\mu^2 \chi_\nu^2 \) is
averaged over one mono-layer. Since the wavefunctions \( \chi(z) \) are envelop functions, they
vary slowly with respect to \( a \), we can therefore drop the average and simply take their
value anywhere in the monolayer centered around position \( z \).

We have to sum all contributions from all monolayers in the spatial region of the alloy. We
can take the limit of the continuum by considering the transformation: \( \sum_{\text{alloy}} a \rightarrow \int_{\text{alloy}} dz \).
5.3. Elastic scattering mechanisms

This transformation is valid only because $\chi(z)$ are slow varying functions. The lattice constant can therefore be considered as a distance element $dz$. This yields:

$$\Gamma_{\text{inter}}^{\nu \rightarrow \mu} = \pi \frac{m_\nu a^3 (\delta E_c)^2 x(1-x)}{\hbar^2} \int_{\text{alloy}} dz \chi_\mu^2(z) \chi_\nu^2(z)$$

The only difference between the above expression and the expression found by T. Unuma [103] is a multiplicative factor of $\pi$. To recover the exact formula, we have to assume that $\Lambda^2 = a^2/\pi$. However, there is a mismatch between this expression and the formula derived in [7,17]. In order to reproduce the results obtained by A. Vasanelli [104], we had to correct the formula with a factor $\frac{1}{8}$. This was without reading carefully T. Unuma’s paper where the exact substitution is given:

$$V_0 \rightarrow \delta E_c \quad \Delta^2 \rightarrow \frac{a^2 x(1-x)}{4} \quad \Lambda^2 \rightarrow \frac{a^2}{2\pi}$$

With these substitutions, we recover the missing $\frac{1}{8}$ factor that was previously recovered from comparison with the work of A. Vasanneli. T. Unuma probably left the expression for alloy disorder with a little algebra error, however all his results were computed with the correct expression. We have:

$$\Gamma_{\text{inter}}^{\nu \rightarrow \mu} = \frac{1}{8} \frac{m_\nu a^3 (\delta E_c)^2 x(1-x)}{\hbar^2} \int_{\text{alloy}} dz \left[ \chi_\mu^2(z) - \chi_\nu^2(z) \right]^2$$

Interface roughness and alloy disorder are fundamentally the same kind of scattering but at a different length scales. Nevertheless we have to evaluate both independently. We do not expect interferences since: $\Lambda_{\text{IFR}} \gg a$.

The scattering caused by alloy disorder does not depend on the exchanged momentum between subbands, since the disorder has an atomic scale. A typical value for the scattering time between subbands separated by about 160 meV (7.5 μm) is ≈ 4.5 ps for lattice matched Ga0.47In0.53As /Al0.48In0.52As. Various lifetimes are summarized in Fig.(5.4).

**Calculation of the $\Gamma_{\text{intra}}$ contribution**

The intra-subband energy broadening due to alloy disorder scattering is given by the limit used in the previous calculations. It does not depend on the exchanged momentum and reads [103]:

$$\Gamma_{\text{intra}}^{(\mu \nu)} = \frac{1}{8} \frac{m^* a^3 (\delta E_c)^2 x(1-x)}{\hbar^2} \int_{\text{alloy}} dz \left[ \chi_\mu^2(z) - \chi_\nu^2(z) \right]^2$$

The latter is easier to evaluate than the energy broadening due to interface roughness. This formula includes only one effective mass $m^*$ for all subbands. We can however generalize
this formula for non-parabolicity. We have:

$$\Gamma_{\text{intra}}^{(\mu\nu)} = \frac{1}{8} a^3 (\delta E_c)^2 \frac{x(1-x)}{\hbar^2} \int_{\text{alloy}} dz \left[ m_\mu \chi_\mu^4(z) + m_\nu \chi_\nu^4(z) - (m_\mu + m_\nu) \chi_\mu^2(z) \chi_\nu^2(z) \right].$$

(5.45)

5.3.3 Ionized impurities scattering

This is the last elastic scattering source we will consider in this work. Quantum cascade structures base on III-V semiconductors alloys are doped with electrons using ionized Si donors. The technique used for the dopants’ implantation is modulatin doping \[18\]. In this technique the dopants are not uniformly distributed inside the structure which would causes a reduced mobility and large optical losses at the laser transition energy, but are concentrated far from the optically active region. The distance between the optically active subbands and the doped region is therefore maximized. In mid-infrared quantum cascade structures, the dopants are typically located in a few layers in the middle of the injector region. The Hartree self-consistent potential can therefore be not negligible as discussed in (4.5). However at the laser transition energy the absorption is reduced. The dopants in the active region contribute positively to the kinetic energy balance as discussed in (7.2). At the same time they reduce the lifetime of the injector states, which enhances the population inversion since the depletion of the lower laser-level is made faster.

The electrons are basically scattered by the Coulomb potential created by the dopants. We write \(N(z)\) the volume concentration of dopants at the position \(z\) in the heterostructure. We also assume that we know the dielectric constant \(\epsilon(z)\).

Calculation of the \(\Gamma_{\text{inter}}\) contribution

The inter-subband scattering between subband \(\mu\) and subband \(\nu\) is given by:

$$\Gamma_{\text{inter}}^{\nu\rightarrow\mu} = \frac{m_\mu q_0^4}{4\pi\epsilon_0\hbar^2} \int_{\text{dopants}} dZ \frac{1}{\epsilon^2(Z)} \int_0^\pi d\theta \times$$

$$\left\{ \frac{1}{\sqrt{q^2 + q_s^2}} \int dz \chi_\mu(z) \chi_\nu(z) \exp \left( -\sqrt{q^2 + q_s^2} |z - Z| \right) \right\}^2$$

(5.46)

where \(\epsilon_0\) is the vacuum permitivity, \(q_0\) the signless elementary charge and \(q_s\) a screening wave-vector. The exchanged wave-vector is the same as for interface roughness scattering in Eq.(5.22) and reads:

$$q^2 = \left( 1 + \frac{m_\mu}{m_\nu} \right) k^2 + \frac{2m_\mu E_{\nu\mu}}{\hbar^2} - 2k \sqrt{\frac{m_\mu}{m_\nu} k^2 + \frac{2m_\mu E_{\nu\mu}}{\hbar^2}} \cos \theta$$

(5.47)
For inter-subband scattering the exchanged wave-vector is usually not zero. But the singularity $q^{-1}$ may cause artificially high scattering rates in the vicinity of $q = 0$. We follow J. H. Smet [91] by considering a screening wave-vector $q_s$. The latter is determined by a simple Debye model as:

$$q_s = \sqrt{\frac{q_0^2 N_s}{\epsilon_0 \langle \epsilon \rangle k_b T_e}}$$

(5.48)

where $N_s$ is the sheet carrier density per period $N_s = L_p^{-1} \int_{\text{period}} dz N(z)$, and $\langle \epsilon \rangle$ is the averaged dielectric constant over one period. The temperature $T_e$ is the electronic temperature as discussed in (7.2).

This screening model is quite poor compared to the self-consistent models developed for far-infrared lasers in Monte-Carlo simulations, to cite only one example [11]. However this minimal model avoid efficiently the $q^{-1}$ divergence and yields reliable numerical results in mid-infrared structures.

**Calculation of the $\Gamma_{\text{intra}}$ contribution**

The intra-subband energy broadening for ionized impurities reads for a pair of subbands $(\mu, \nu)$ with $m_\mu = m_\nu = m^*$:

$$\Gamma_{\text{intra}}^{(\mu, \nu)} = \frac{m^* q_0^4}{4\pi \epsilon_0 h^2} \int_{\text{dopants}} dZ N(Z) \frac{1}{\epsilon^2(Z)} \int_0^\pi d\theta \times$$

$$\left\{ \frac{1}{\sqrt{q^2 + q_s^2}} \int dz \left[ \chi_\mu^2(z) - \chi_\nu^2(z) \right] \exp \left( - \sqrt{q^2 + q_s^2} |z - Z| \right) \right\}^2$$

(5.49)

where the exchanged momentum $q$ reads as in Eq.(5.26):

$$q^2 = 2k^2 (1 - \cos \theta).$$

(5.50)

The inclusion of non-parabolicity effects with different masses for subband $\mu$ and subband $\nu$ is less simple than in the previous cases. First we have to expand the previous integral. In order to lighten the notation we set $Q = \sqrt{q^2 + q_s^2}$ and we write the exponential kernel as: $\exp \left( - Q |z - Z| \right) \equiv K_Q(z, Z)$. With these notations, we have:

$$\left[ \frac{1}{Q} \int dz \left[ \chi_\mu^2(z) - \chi_\nu^2(z) \right] K_Q(z, Z) \right]^2 = \left[ \frac{1}{Q} \int dz \chi_\mu^2(z) K_Q(z, Z) \right]^2$$

$$+ \left[ \frac{1}{Q} \int dz \chi_\nu^2(z) K_Q(z, Z) \right]^2 - \frac{2}{Q^2} \int dz dz' \chi_\mu^2(z') \chi_\nu^2(z) K_Q(z, Z) K_Q(z', Z) K_Q(z, Z)$$

(5.51)
We have to consider the independent integrals:

\[ \Phi_\mu(Q, Z) \equiv \frac{1}{Q} \int dz \chi_\mu^2(z) K_Q(z, Z) \] (5.52)

The above expansion is therefore rewritten in terms of \( \Phi \) by:

\[
\left[ \frac{1}{Q} \int dz \left[ \chi_\mu^2(z) - \chi_\nu^2(z) \right] K_Q(z, Z) \right]^2 = \Phi_\mu^2(Q, Z) + \Phi_\nu^2(Q, Z) - 2\Phi_\mu(Q, Z)\Phi_\nu(Q, Z) \] (5.53)

We can therefore rewrite the intra-subband broadening as:

\[
\Gamma_{\text{intra}}^{(\mu, \nu)} = \frac{m^* q_0^4}{4\pi\epsilon_0 \hbar^2} \int_{\text{dopants}} dZ \frac{N(Z)}{e^2(Z)} \int_0^\pi d\theta \times \left\{ m_\mu \Phi_\mu^2(Q, Z) + m_\nu \Phi_\nu^2(Q, Z) - (m_\mu + m_\nu) \Phi_\mu(Q, Z)\Phi_\nu(Q, Z) \right\}^2 \] (5.54)

This generalized expression have to be evaluated carefully, at least if we keep an arbitrary shape for the dopants profile \( N(z) \).

**Computation of the ionized impurities form factors**

The computation of the ionized impurities scattering is in general not a light numerical task if we assume an arbitrary profile of dopants \( N(z) \). Its optimization was however crucial in order to obtain a numerically efficient transport model.

The basic idea is to split the form factor into simple integrals defined like \( \Phi_\mu(Q, Z) \) previously (but without the factor \( Q^{-1} \)). We then exploit the exponential kernel \( K_Q(z, Z) \). It is clear that changing of coordinates to \( D = z - Z \) and \( S = z + Z \) allows to compute the exponential kernel up to a cut-off value. However, the basic integrals have to be evaluated for all wavefunctions and all values of \( q \), which is numerically very heavy.

In order to reduce the computational time, we have used a cubic interpolation, by computing simultaneously \( Q^{-1}\Phi_\mu(Q, Z) \) and its first derivative with respect to \( q \). We have imposed the value of the function and of its first derivative on a cubic polynomial. Due to the smoothness of the form factor we were able to dynamically fit the form-factor using very few exact values computed on a special mesh of the \( q \)-axis. The latter was defined by choosing an interval defined by \( q_{\text{min}} \) and \( q_{\text{max}} \) and harmonic intermediate points. For example between \( q = 0 \) and \( q = q_{\text{max}} \) we have intermediate points defined by \( q_{\text{max}}/2, q_{\text{max}}/4, \ldots, q_{\text{max}}/2^n \) where \( n \) is the level-of-details of the form-factor. This mesh is very efficient as the exponential rapidly dominates the form-factor. Therefore the highest resolution is only required for small values of \( q \).

We have implemented a similar algorithm for the form-factor of the LO-phonon interaction. The latter is described below with more details (5.4).
5.4 Inelastic non-radiative scattering mechanisms: LO-phonon scattering

In this section we examine the most efficient inelastic scattering mechanism for mid-infrared quantum cascade lasers based on III-V alloys. The latter is caused by scattering of polar longitudinal-optical phonons as shown by the seminal work of P. J. Price [78]. We do not want to reproduce all the calculations here, however in polar semiconductors (like III-V alloys), the interaction between an electron and the lattice is described by the Fröhlich Hamiltonian [34]. The latter models the motion of a conduction electron in an ionic material. The motion of the electron in the polar material causes a deformation of the lattice. The local and partially unscreened ionic potential (piezo-electric potential) interacting back with the electron. This gives a self-energy to the latter. The electron and the phonons can therefore be considered as a quasi-particle: the polaron, where the electron is dressed by polar optical phonons.

Here we will only consider the Fröhlich Hamiltonian (that describe longitudinal polar optical phonons) and calculate first order scattering rates with electrons. Of course the phonon has an energy dispersion given by $\hbar \omega(q)$ where $q$ is the phonon wave-vector. We assume that this dispersion is negligible here (this is the case for most III-V alloys) and that we can estimate the energy of the LO-phonon by its long wavelength value: $\hbar \omega_{LO} \equiv \hbar \omega_{LO}(q = 0)$.

The phonon interaction is further simplified by considering only bulk phonon modes of the well-material. However the spatial variation of the dielectric constant in the heterostructure gives rise to confined phonon modes [42]: phonon can be confined inside a material layer, while others are confined at the interfaces between the materials. The strength of the confinement however depends on the contrast of the dielectric constant. The latter is not so small for $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ ($\epsilon_{\infty} \approx 10.92$) and $\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ ($\epsilon_{\infty} \approx 8.16$) lattice matched on InP. However this topic was rather addressed in the GaAs/AlAs material system. Nevertheless a sum-rule argument [42] shows that the total contributions of squared matrix elements of all phonon modes should be equal to the bulk approximation. This comfort us in the use, as a first approximation, of bulk LO-phonons.

Here we do not give details of the calculations of the scattering rates and intra-subband energy broadening. The transition rate between electronic states is computed using Fermi’s Golden Rule and the Fröhlich Hamiltonian. An important fact is that the LO-phonon in a heterostructure has approximatively the same properties as a three-dimensional (bulk) LO-phonon (the confinement being neglected). The conservation of the momentum in the $z$-direction in three-dimensional system gives the two-dimensional
system matrix elements [103]:

$$|M_{2D}|^2 = \sum_{q_z} |M_{3D} I_{\mu\nu}(q_z)|^2.$$  \hspace{1cm} (5.55)

Where the form-factor $I_{\mu\nu}(q_z)$ between two subbands $\mu$ and $\nu$ is given by:

$$I_{\mu\nu}(q_z) = \int dz \chi_\mu(z) \chi_\nu(z) e^{i q_z z} \hspace{1cm} (5.56)$$

The three-dimensional matrix elements for emission and absorption of bulk phonons are given by:

$$\langle |M_{3D}|^2 \rangle = \frac{q_0^2 \hbar \omega_{\text{LO}} (N_{\text{LO}} + 1)}{2 \varepsilon_0 Q^2} \left( \frac{1}{\varepsilon_{\infty}} - \frac{1}{\varepsilon_0} \right) \text{ for emission}$$

$$\langle |M_{3D}|^2 \rangle = \frac{q_0^2 \hbar \omega_{\text{LO}} N_{\text{LO}}}{2 \varepsilon_0 Q^2} \left( \frac{1}{\varepsilon_{\infty}} - \frac{1}{\varepsilon_0} \right) \text{ for absorption}$$  \hspace{1cm} (5.57)

where $Q$ is the absolute value of the three-dimensional scattering vector, $\varepsilon_0$ the static dielectric constant and $\varepsilon_{\infty}$ the high-frequency dielectric constant and $\varepsilon_0$ the vacuum permittivity.

The occupancy function for bosons being given by the Bose-Einstein distribution:

$$N_{\text{LO}} = \frac{1}{\exp \left( \frac{\hbar \omega_{\text{LO}}}{k_b T} \right) - 1} \hspace{1cm} (5.58)$$

It is worth to say that the temperature of the phonon-bath $T$ is usually considered as the lattice temperature. The latter is given by the averaged temperature of the active region in a real device.

In this model we neglect the hot-phonons effects [58,67,76,93].

**Calculation of the $\Gamma_{\text{inter}}$ contribution**

The Eq.(5.1) needs to be evaluated with the correct dirac function (5.1.1) for both emission and absorption processes. The inter-subband energy broadening is given by:

$$\Gamma_{\text{inter}}^{(\nu \rightarrow \mu)}(k) = \frac{m_{\mu} q_0^2 \hbar \omega_{\text{LO}}}{4 \pi \varepsilon_0 \hbar^2} \left( \frac{1}{\varepsilon_{\infty}} - \frac{1}{\varepsilon_0} \right) \int_0^\pi d\theta \times$$

$$\left[ \Theta(\epsilon_{\nu}(k) + E_{\nu \mu} - \hbar \omega_{\text{LO}})(N_{\text{LO}} + 1) \Phi_{(\nu \mu), (\nu \mu)}(\sqrt{q_e^2 + q_s^2}) \right.$$

$$\left. + \Theta(\epsilon_{\nu}(k) + E_{\nu \mu} + \hbar \omega_{\text{LO}})(N_{\text{LO}}) \Phi_{(\nu \mu), (\nu \mu)}(\sqrt{q_e^2 + q_s^2}) \right] \hspace{1cm} (5.59)$$
5.4. Inelastic non-radiative scattering mechanisms: LO-phonon scattering

where \( \Theta(x) \) is the Heaviside step function: \( \Theta(x) = 0 \) for \( x \leq 0 \) and \( \Theta(x) = 0 \) for \( x > 0 \), \( E_{\nu\mu} = \varepsilon_{\nu}(0) - \varepsilon_{\mu}(0) \) is the detuning energy between the subband edges and \( \varepsilon_0 \) is the vacuum permittivity. The Heaviside functions were introduced to underline the energy conservation in the emission and absorption processes.

We are considering screened wave-vectors with \( \sqrt{q_{c,a}^2 + q_s^2} \), the screening length being given by a Debye model as in Eq.(5.48). If the interaction is unscreened, phonon emission/absorption occurring near (or at) a resonance \( \Delta_{\nu\mu} = \hbar \omega_{\text{LO}} \) will have a divergent (\( \sim q^{-1} \)) form-factor.

The energy conservation yields an exchanged momentum \( q = k - k' \) for the emission \( q_e \) and for the absorption \( q_a \). Both are given below:

\[
q_e^2 = \left(1 + \frac{m_\mu}{m_\nu} \right) k^2 + \frac{2m_\mu}{\hbar^2} (E_{\nu\mu} - \hbar \omega_{\text{LO}}) - 2k \sqrt{\frac{m_\mu}{m_\nu} k^2 + \frac{2m_\mu}{\hbar^2} (E_{\nu\mu} - \hbar \omega_{\text{LO}}) \cos \theta}
\]

The exchanged momentum in the absorption process is given by the transformation \( \hbar \omega_{\text{LO}} \rightarrow -\hbar \omega_{\text{LO}} \). We have:

\[
q_a^2 \equiv q_e^2 (\hbar \omega_{\text{LO}} \rightarrow -\hbar \omega_{\text{LO}})
\]

The form factor \( \Phi_{(\nu\mu)(\mu\nu)}(q) \) that appears in Eq.(5.59) is defined by:

\[
\Phi_{(\nu\mu)(\mu\nu)}(q) \equiv \frac{1}{q} F_{(\nu\mu)(\mu\nu)}(q).
\]

The form-factor \( F \) is defined by:

\[
F_{(\mu\nu)(\sigma\tau)}(q) \equiv \int dz \int dz' \chi_\mu(z) \chi_\nu(z) \chi_\sigma(z') \chi_\tau(z') \exp (-q |z - z'|)
\]

Where we have assumed that the wavefunctions are real. For bound-states in a one-dimensional problem this choice is always possible. The form-factor \( F \) can be computed very efficiently by dynamical fitting methods on a special mesh of the \( q \)-axis (5.4).

**Calculation of the \( \Gamma_{\text{intra}} \) contribution**

The energy broadening has two parts: one for the emission process, the other for the absorption. The formula that includes the non-parabolicity does not present special diffi-
But for the Dirac function $\Phi$ case for the emission process as the threshold energy is to the intra-subband broadening as such transitions are always possible. This is not the case for the absorption process as the threshold energy is $\hbar \omega_{LO}$. The form-factors $\Phi^{(\mu\mu),(\nu\nu)}$ and $\Phi^{(\nu\nu),(\mu\mu)} = \Phi^{(\mu\mu),(\nu\nu)}$ are defined in Eq. (5.62).

The exchange momentum for the emission/absorption processes depends on the Dirac functions: $\delta (\epsilon_\mu (k) - \epsilon_\nu (k') \pm \hbar \omega_{LO})$. We have:

$$q^{2}_{e,\mu} = 2k^2 - \frac{2m_\mu \hbar \omega_{LO}}{\hbar^2} - 2k \sqrt{k^2 - \frac{2m_\mu \hbar \omega_{LO}}{\hbar^2}} \cos \theta \quad \text{and} \quad q^{2}_{a,\mu} \equiv q^{2}_{e,\mu} (\hbar \omega_{LO} \rightarrow -\hbar \omega_{LO}). \quad (5.65)$$

But for the Dirac function $\delta (\epsilon_\nu (k) - \epsilon_\nu (k') \pm \hbar \omega_{LO})$, we have:

$$q^{2}_{e,\nu} = 2k^2 - \frac{2m_\nu \hbar \omega_{LO}}{\hbar^2} - 2k \sqrt{k^2 - \frac{2m_\nu \hbar \omega_{LO}}{\hbar^2}} \cos \theta \quad \text{and} \quad q^{2}_{a,\nu} \equiv q^{2}_{e,\nu} (\hbar \omega_{LO} \rightarrow -\hbar \omega_{LO}). \quad (5.66)$$

From the above equations it is clear that we need to optimize the computation of the form-factors in a numerical implementation.

**Computation of the form factor $F^{(\mu\nu)(\sigma\tau)}(q)$**

The form-factor $F$ is defined in Eq. (5.63):

$$F^{(\mu\nu)(\sigma\tau)}(q) \equiv \int dz \int d\tilde{z}' \chi_\mu (z) \chi_\nu (z) \chi_\sigma (\tilde{z}') \chi_\tau (\tilde{z}') \exp \left\{ -q |z - \tilde{z}'| \right\} \quad (5.67)$$

For the computation of the LO-phonon transition rates we need to consider various configurations of the indices: $(\nu\mu)(\mu\nu)$, $(\nu\nu)(\nu\nu)$, $(\mu\mu)(\nu\nu)$, and $(\nu\nu)(\mu\mu)$. The computation is
5.5 Typical scattering rates

We briefly show some computations of inter-subband scattering rates. We consider a square quantum well made of GaInAs/AlInAs. The transition from the first excited state at \( \mathbf{k} = 0 \) to the ground-state is considered for all the interactions we have implemented in this work. To extend the scope of validity of the computation, we have considered a squared well of numerically heavy when we have to evaluate this form-factors for a large number of values of \( q \) and for several subbands.

We have implemented a routine that allows the accurate evaluation of the form-factor with a very limited number exact evaluations. The form-factor is dominated by the exponential, except for the \( q = 0 \) case. However this never occurs since we are usually evaluating \( F \) for screened wave-vectors. Anyway, this case is treated by the numerical mode.

We perform a change of coordinates to \( x = z - z' \) and \( y = z + z' \). The exponential kernel is therefore represented by \( e^{-q|x|} \). As the integrant is a product of wavefunctions, we expect that the form-factor will have a complicated dependence for small values of \( q \). After some analysis, we decided to interpolate it with a cubic polynomial. The choice of the mesh points of the \( q \)-axis, where the form-factor is computed exactly, was determined by the requirement of having the maximal resolution for small values.

We first compute the boundary values \( q_{\text{min}} \) and \( q_{\text{max}} \), and we consider the mesh points:

\[
\{q_k\} \equiv \{q_{\text{min}}, q_{\text{min}} + \frac{\Delta q}{2^n}, q_{\text{min}} + \frac{\Delta q}{2^{n-1}}, \ldots, q_{\text{min}} + \frac{\Delta q}{2}, q_{\text{max}}\}, \tag{5.68}
\]

where \( \Delta q \equiv q_{\text{max}} - q_{\text{min}} \) and \( n \) is an integer (the level-of-details). For mid-infrared quantum cascade structures, this number is typically 8 (and can be kept for far-infrared structures).

We have to assume that the form-factor includes a screening vector by letting \( \tilde{F}(q) \equiv F(\sqrt{q^2 + q_s^2}) \). This is necessary for the following reason. At each point \( q_k \), the form-factor and its first derivative with respect to \( q \) can be evaluated simultaneously. If the screening wave-vector was not included, the evaluation of the first derivative would be wrong.

We then perform interpolation between the adjacent points \( q_k \) and \( q_{k+1} \) with a cubic polynomial. We impose at the two boundaries the value of the form-factor and of its first derivative. This yields an accurate estimation of the form-factor.

In Fig.(5.3) the form-factors \( F_{(01)(10)}(q) \) and \( F_{(00)(11)}(q) \) are shown for the ground-state (0) and the first-excited state (1) of a single-quantum-well active region shown in Fig.(5.2a). A level-of-details of \( n = 8 \) is a secure value.
Figure 5.3: Two form-factors, found in transition rates due to LO-phonon interaction, are shown. The dynamical fitting method described in (5.4) is applied for some sampling values \( n = 2, 3, 4, \infty \). The interpolation is very efficient even for low \( n \) values. The structure considered in this illustration is described in the text.

lattice matched \( \text{Ga}_{0.47}\text{In}_{0.53}\text{As} / \text{Al}_{0.48}\text{In}_{0.52}\text{As} \) and one with strain-compensated materials: \( \text{Ga}_{0.36}\text{In}_{0.64}\text{As} / \text{Al}_{0.67}\text{In}_{0.33}\text{As} \). The results are shown in Fig.(5.4).

The parameters for the interface-roughness-scattering are \( \Delta = 1.2 \text{ Å} \), \( \Lambda = 9.0 \text{ nm} \) and \( \kappa = 1.5 \text{ nm} \). For the alloy disorder the fraction are given above and the lattice constant is \( a \approx 5.65 \text{ Å} \). The conduction band discontinuity is \( \delta E_c \approx 0.52 \text{ eV} \) for lattice matched and \( \delta E_c \approx 0.8 \text{ eV} \) for strain-compensated material. The energy of LO-phonon is \( \hbar \omega_{\text{LO}} = 34 \text{ meV} \).

These results show that even in the strain-compensated materials the alloy disorder does not dominates the elastic scattering rates. The crossing between elastic and inelastic processes shows that for a larger subband separation the phonon interaction becomes less efficient as the exchanged momentum increases. The behavior of the interface roughness scattering is more complicated. We have seen in Eq.(5.25) that the form-factor of the interface roughness scattering decay exponentially with the separation energy. However, this reduction is compensated by a reduction of the well width. The value of the wavefunctions at the interfaces is therefore larger and the coupling constant \( F^2 \) in Eq.(5.20) increases.
5.6 Other sources of scattering

Here we have neglected two other main sources of scattering: the longitudinal acoustical phonons (LA-phonons) and the electron-electron scattering, although the latter is treated in the mean-field approximation, through a self-consistent Hartree potential (4.5).

Regarding the LA-phonon interaction [33], the latter was here initially considered as a virtually elastic scattering [103], since the acoustic phonons’ energy-dispersion is linear for small $|Q|$ values: $\hbar \omega(Q) = c_s \left( q_\parallel^2 + q^2 \right)$, where $c_s$ is the longitudinal speed of sound. The momentum parallel to the heterostructure direction is $q_\parallel$ and $q$ is the exchanged momentum in inter- or intra-subband scattering, selected by the dirac function $\delta(\varepsilon_\nu(k) - \varepsilon_\mu(k') \pm \hbar \omega(q))$ for the inter-subband case. The sign ($-$) meaning emission and the sign ($+$) absorption. The coupling constant of bulk acoustical phonons is proportional to the thermal energy $k_b T$ in the elastic approximation. We therefore expect that this interaction
becomes relevant with temperature. However to have a sufficient accuracy it is necessary to implement this interaction as an inelastic process. Nevertheless by examining the typical lifetimes computed in [33, 91, 104], we do not expect this scattering mechanism to largely alter the transport (especially the upper laser-state lifetime) in mid-infrared quantum cascade structures. However in a future development of the transport model presented in (7), this interaction should be fully taken into account.

The other important missing interaction is the electron-electron interaction. In general the latter is quite complicated [91] to compute and therefore to implement practically with the same generality as the previously examined interactions. The latter is an elastic scattering mechanism that involves two electrons. It tends to redistribute electrons in subbands (we assume two subbands labeled by 2 and 1, where \( E_{21} > 0 \)), with \( 2, 1 \rightarrow 2, 1 \) and \( 2, 2 \rightarrow 2, 2 \) transitions, and reduces the upper-state population, with \( 2, 2 \rightarrow 1, 1 \) and \( 2, 2 \rightarrow 2, 1 \) transitions. The effect on the lifetime of subband seems to be negligible for mid-infrared quantum cascade structures. However we have implemented this interaction implicitly by assuming intra-subband thermal distributions (7.2). The redistribution of electrons in a thermal distribution in subbands being mainly achieved by electron-electron scattering. It is worth to mention that this interaction may not be neglected in far-infrared structures [91]. As for the LA-phonon interaction, electron-electron interaction should be implemented in a future development of the transport model presented in this work.
Chapter 6

Basic transport models and requirements

This is a central part of this work. In this section we first review the rate equations in the Boltzmann equations (6.1). We examine various level of approximation that will be useful later for the transport modeling.

The usual rate equations in quantum cascade structures (6.2) are then discussed. The latter are used in a more general perspective by taking into account three fundamental periods, allowing the definition of a current (6.3). The latter is found unstable when resonances arise between subbands. This motivates the investigation of sequential resonant tunneling across a barrier (6.4). In this part density matrix models for 0-dimensional states are investigated (6.4.3). The periodicity of the structure is then implemented on simple models (6.4.5). Finally, a toy model with four 0-dimensional states is formulated and applied to a far-infrared single-quantum-well structure. The scattering times are adjusted to reproduce the measurements. The features predicted by the model were observed on experimental data (6.5).

The 0-dimensional systems fail to model correctly the transport between subbands. This is shown on two relevant examples: the current density between ground-states in a superlattice (6.6.1) and the leakage current from the upper laser-subband in a mid-infrared quantum cascade structure (6.6.1). This requires a better modeling of the current density between subbands.

The minimal model, a pair of subbands in then solved in the density-matrix formalism (6.6). The calculations are done up to second-order terms. This yields a current that addresses the previously studied issues (6.6.4-6.6.4).
6.1 Derivation and general aspects of rate equations

We examine the simplest semiclassical model that consists of rate equations between the energy-eigenstates inside the period of a quantum cascade heterostructure.

This model is semi-classical: the quantum ingredients are the band structure and the scattering rates (when they are computed and not estimated). The structure of the transport model itself has the semi-classical form of a simplified Boltzmann equation.

In the first sections we examine the structure of rate equations with different levels of approximation (6.1.1-6.1.3). We then give general solutions in relevant configurations (6.1.4-6.1.6).

6.1.1 Rate equations from the Boltzmann equations in a superlattice

We start from the expression of the Boltzmann equation in a superlattice under an applied electric field $F$ (modeling the miniband transport). It reads [110]:

$$\frac{\partial f(q,k,\nu,t)}{\partial t} + \frac{eF}{\hbar} \frac{\partial f}{\partial q} = \left( \frac{\partial f}{\partial t} \right)_{\text{scatt}}$$ (6.1)

where $q$ is the crystal momentum in the superlattice direction, $k$ is the in-plane momentum and $\nu$ is the miniband index. The scattering term is given by:

$$\left( \frac{\partial f}{\partial t} \right)_{\text{scatt}} = \sum'_{q',k',\nu'} \int_{-\pi/d}^{\pi/d} dq' \int d^2k' \times \left[ P(q',k',\nu' \rightarrow q,k,\nu) f(q',k',\nu',t)(1 - f(q,k,\nu,t)) - P(q,k,\nu \rightarrow q',k',\nu') f(q,k,\nu,t)(1 - f(q',k',\nu',t)) \right].$$ (6.2)

The functions $P(q,k,\nu \rightarrow q',k',\nu')$ give the scattering probabilities from state $q,k,\nu$ to state $q',k',\nu'$. The occupation function $f(q,k,\nu,t)$ being the solution to the problem. The latter is usually computed with a Monte-Carlo method [51]. Here $q$ is the Bloch wave-vector in the first Brillouin mini-zone, $k$ is the in-plane wave-vector and $\nu$ is the miniband index.

The equations we develop here are pretty much simpler, but they obey to the same physical model. In a localized period of a quantum cascade structure, the quantization gives subbands indexed by an integer $\nu$ (here we keep the same notation as for the miniband index) with an in-plane wave-vector $k$. The $q$ wave-vector is therefore meaningless. In particular the acceleration term $(eF/\hbar) \partial/\partial q$ vanish.
If we integrate \( f(k, \nu, t) \) over the in-plane wave-vectors with a density of states \( D_k \), we can define the total population in subband \( \nu \) by:

\[
N_\nu(t) = D_k \int d^2 k f(k, \nu, t) \tag{6.3}
\]

If now we apply integration on both side of the Eq.(6.1) with \( D_k \int d^2 k \), we get:

\[
\frac{dN_\nu(t)}{dt} = \sum_{\nu'} \left( W_{\nu' \rightarrow \nu}[f_{\nu'}, f_\nu] - W_{\nu \rightarrow \nu'}[f_\nu, f_{\nu'}] \right). \tag{6.4}
\]

where we have modified the notation of the occupancy function \( f(k, \nu, t) \rightarrow f_\nu(k, t) \). We also have introduced the functional:

\[
W_{\nu \rightarrow \nu'}[f_\nu, f_{\nu'}] = \int d^2 k' D_k \int d^2 k P(k\nu \rightarrow k'\nu') f_\nu(k, t)(1 - f_{\nu'}(k', t)) \tag{6.5}
\]

The term \( P(k\nu \rightarrow k'\nu') f_\nu(k, t) \) gives the fraction of the carriers that performs the transition \( k\nu \rightarrow k'\nu' \). The multiplication by \((1 - f_{\nu'}(k', t))\) reduces it eventually if the final state \( k'\nu' \) is occupied. An integration over every initial state \( k \) in the subband \( \nu \) is performed and multiplied by the density of states \( D_k \). This gives the number of carriers that are transferred from subband \( \nu \) into the final state \( k'\nu' \) of subband \( \nu' \). The final integration runs over all the final states \( k' \). The term \( W_{\nu \rightarrow \nu'} \) therefore gives the number of electrons performing a transition from subband \( \nu \) to subband \( \nu' \).

The functional notation needs to be kept as the internal definition of \( W_{\nu \rightarrow \nu'} \) require the knowledge of the distribution \( f_\nu(k, t) \) and \( f_{\nu'}(k, t) \).

The transition probabilities need to be computed between arbitrary subband-states \( k\nu \) and \( k'\nu' \). This can be achieved by using the Fermi’s Golden Rule for example. The major complication is that the functions \( f_\nu(k, t) \) need to be computed, which is in turn the utility of the previous equations.

By now we will perform approximations and obtain step after step rates equations models with a decreasing level of details. This enables to compute accurate results in a simpler manner.

### 6.1.2 Subbands with thermal distributions

The first interesting step is to reduce the knowledge that we have of the intra-subband distribution of carriers \( f_\nu(k, t) \). The ability to describe non-equilibrium distributions in subbands would be very relevant for a precise modeling of the dynamics of electrons in heterostructures and may lead to the discovery of new effects. Here we are going the way back by simplifying the problem.
We simply the functions $f_\nu$ by considering thermal distributions with a Fermi-Dirac model. As we impose thermal distributions, it is meaningless of looking at the intra-subband dynamics we therefore set for every subbands:

$$P(k_\nu \rightarrow k_\nu') \equiv 0 \Rightarrow W_{\nu\nu} \equiv 0$$  \hspace{1cm} (6.6)

This restriction is only conceptual as in Eq.(6.4), the $\nu \rightarrow \nu$ contributions cancel. We then replace the general distributions by Fermi-Dirac distributions:

$$f_\nu(k, t) \rightarrow f_{FD}(E_\nu(k), \mu_\nu(t), \beta_\nu).$$

It is worth to remark that the Fermi-level depends on time and that we kept a different temperature $T_\nu$ for each subbands with $\beta_\nu = 1/k_b T_\nu$, where $k_b$ is the Boltzmann constant.

We have:

$$f_{FD}(E_\nu(k), \mu(t), \beta_\nu) = \frac{1}{1 + \exp[\beta_\nu(E_\nu(k) - \mu_\nu(t))]} \text{ with } E_\nu(k) = E_{0\nu} + \frac{\hbar^2 k^2}{2m_\nu}.$$  \hspace{1cm} (6.7)

The subband dispersion is calculated with a first order perturbation as in Eq.(3.25), with an averaged mass $m_\nu \equiv m_\nu$ (for comoditiy). The subband-edge energy is $E_{0\nu}$.

This distribution can be integrated over the energy directly as the energy dispersion depends on $k^2$. It yields:

$$N_\nu(t) = \frac{D_0 m_\nu}{\beta_\nu} \log \left[1 + \exp (\beta_\nu (\mu_\nu(t) - E_{0\nu}))\right] \text{ where } D_0 = \frac{1}{\pi \hbar^2} \hspace{1cm} (6.8)

The time dependence of the Fermi-level gives the time dependence of the total population of the subband $N_\nu(t)$.

The functional $W_{\nu\rightarrow\nu'}$ can be rewritten according to these simplifications. It becomes a functional of the total population of subband $\nu$ and $\nu'$:

$$W_{\nu\rightarrow\nu'}[N_\nu, N_{\nu'}] = \int d^2 k' \int d^2 k P(k_\nu \rightarrow k_\nu') f_{\nu}(E_{\nu}(k), N_{\nu}(t), \beta_\nu) \times \left[1 - f_{\nu'}^{FD}(E_{\nu'}(k), N_{\nu'}(t), \beta_\nu)\right]$$

Even with thermal distributions in subbands the Eq.(6.4) is a complicated non-linear differential system. This provides averaged transitions rates $W_{\nu\rightarrow\nu'}[N_\nu, N_{\nu'}]$. These expressions will be used in the effective transport model we develop in (7.1). A self-consistency loop being used to converge on the solution.

The averaged transition rate $\overline{w}_{\nu\rightarrow\nu'}$ between subband $\nu$ and $\nu'$ is obtained with the following definition:

$$\overline{w}_{\nu\rightarrow\nu'} \equiv \frac{1}{N_\nu(t)} W_{\nu\rightarrow\nu'}[N_\nu, N_{\nu'}].$$

(6.10)

This rate is then inserted in very simple rate equations that we will obtain below, when the population and probability transition can be factorized. Since this method is self-consistent it yields accurate results through simple algebraic equations in steady-state.
Low electron densities

For low electron densities Eq.(6.9) can be simplified by removing the occupancy of the final subband $\nu$:

$$W_{\nu \rightarrow \nu'}[N_{\nu}] = \int d^2k' D_k \int d^2k P(k\nu \rightarrow k'\nu') f_{\nu}(E_{\nu}(k), N_{\nu}(t), \beta_{\nu})$$

(6.11)

which in turn remove the functional decency of $W_{\nu \rightarrow \nu'}$ on $N_{\nu'}$. Averaged transition rates can also be computed. However we will keep the occupation factor of the final subband when we will treat effective transport models (7).

6.1.3 Factorization of the distribution

The final approximation we may discuss yields the standard rate equations. For that we do not need to choose a particular thermal distribution and we may keep the general $f_{\nu}(k, t)$. Nevertheless for the consistency of the model, we have kept the assumption $W_{\nu\nu'} \equiv 0$ although the dynamics of the populations is not affected by it.

This approximation relies on the factorization of the transition probability $P(k\nu \rightarrow k'\nu')$ from the carrier distributions. This approximation is meaningless is we keep the $(1 - f_{\nu'}(k', t))$ occupation factor. We write:

$$P(k\nu \rightarrow k'\nu') = w_{0}^{\nu \rightarrow \nu'} \delta(k')$$

(6.12)

Where $w_{0}^{\nu \rightarrow \nu'}$ is a constant transition rate that we can evaluate anywhere between the initial and the final subbands. It is usually evaluated for $k = 0$ in the initial subband: $w_{0}^{\nu \rightarrow \nu'} = P(0\nu \rightarrow k'\nu')$. We have kept $k$ for the final state as even if we choose $k = 0$ (single point), the final states wave-vectors $\{k'\nu\}$ can be a point or a circle in the lower subband. In some models it will be also useful to compute this lifetime at a characteristic thermal energy $k_{B}T_0$ in the initial subband.

We insert this expression in Eq.(6.5) without the final occupation factor and we get:

$$W_{\nu \rightarrow \nu'}[f_{\nu}, f_{\nu'}] = \int d^2k' D_k \int d^2k P(k\nu \rightarrow k'\nu') f_{\nu}(k, t)$$

$$= w_{0}^{\nu \rightarrow \nu'} D_k \int d^2k f_{\nu}(k, t)$$

$$= w_{0}^{\nu \rightarrow \nu'} N_{\nu}(t).$$

(6.13)

We can rewrite Eq.(6.4) in a conventional form:

$$\frac{dN_{\nu}(t)}{dt} = \sum_{\nu'} (w_{\nu \rightarrow \nu'}^{0} N_{\nu}(t) - w_{\nu' \rightarrow \nu}^{0} N_{\nu'}(t))$$

$$= -N_{\nu}(t) \left( \sum_{\nu' \neq \nu} w_{\nu \rightarrow \nu'}^{0} \right) + \sum_{\nu'} w_{\nu' \rightarrow \nu}^{0} N_{\nu'}(t)$$

(6.14)
By introducing the lifetime $\tau_{\nu'\to\nu}$ as being the inverse of the scattering rate $w^0_{\nu'\to\nu}$, we obtain:

$$\frac{dN_{\nu}(t)}{dt} = -\frac{N_{\nu}(t)}{\tau_{\nu}} + \sum_{\nu'} \frac{N_{\nu'}(t)}{\tau_{\nu'\to\nu}}.$$  \hspace{1cm} (6.15)

The total lifetime of subband $\nu$ was introduced $\tau_{-1} = \sum_{\nu' \neq \nu} \tau_{\nu'\to\nu}^{-1}$. This is the transition from subband $\nu$ to all other subbands. We have noticed before in (6.1.2) that we may use averaged scattering rates in the above equation. This is achieved with the substitutions $w^0_{\nu'\to\nu} \rightarrow \bar{w}_{\nu'\to\nu}$.

### 6.1.4 Isolated subbands: matrix formulation and solution

The rate equations can be written in matrix form. If we write $\mathbf{N}(t)$ the vector of total populations and $\mathbf{W}$ a matrix that contains the lifetimes $\tau_{\nu'\to\nu}$, we have:

$$\dot{\mathbf{N}} = \mathbf{W} \cdot \mathbf{N}$$ \hspace{1cm} (6.16)

where we have introduced the notation $\dot{\mathbf{N}} = d\mathbf{N}/dt$. If we assume that the problem involves $N$ subbands we have:

$$\mathbf{W} = \begin{pmatrix} -w_1 & w_{2\rightarrow1} & \ldots & \ldots & \ldots & \ldots & \ldots & w_{N\rightarrow1} \\ w_{1\rightarrow2} & -w_2 & w_{3\rightarrow2} & \ldots & \ldots & \ldots & \ldots & w_{N\rightarrow2} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ w_{1\rightarrow k} & \ldots & w_{k-1\rightarrow k} & -w_k & w_{k+1\rightarrow k} & \ldots & \ldots & w_{N\rightarrow k} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ w_{1\rightarrow N} & \ldots & \ldots & \ldots & w_{N-2\rightarrow N} & w_{N-1\rightarrow N} & -w_N \end{pmatrix}. \hspace{1cm} (6.17)

We have used the notation $w = \tau^{-1}$ in order to enhance the readability. This homogeneous problem can be solved directly in the diagonal basis. Let’s assume that the matrix $U$ makes the matrix $\mathbf{W}$ diagonal with: $\Lambda = U \cdot \mathbf{W} \cdot U^{-1}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ is diagonal. Letting $\mathbf{v} \equiv U \cdot \mathbf{N}$, the differential equation reads in the diagonal basis: $d\mathbf{v}/dt = \Lambda \cdot \mathbf{v}$. The solution is $\mathbf{v}(t) = \exp(\Lambda t) \cdot \mathbf{v}(0)$, with the initial vector $\mathbf{v}(0)$. In the initial basis we therefore have: $\mathbf{N}(t) = U^{-1} \cdot \exp(\Lambda t) \cdot U \cdot \mathbf{N}(0)$. The matrix $e^M = \left( \exp(\lambda_k t) \right)$ is diagonal as $\Lambda$ is diagonal.

With this method the time evolution of the subband populations is known for a particular initial distribution $\mathbf{N}(0)$. It is worth to remark that the system of $N$ subbands is isolated: there is no external driving term. If we compute all the scattering rates between the subbands by using a LO-phonon interaction, the solution will converge to a thermal distribution across the subbands, the temperature being given by the phonon-bath temperature.
This stationary solution is obtained in the long-time limit: \( t \to \infty \). If we apply this limit to the general homogeneous solution, the result will depend on the eigen-values \( \lambda_k \) of the matrix \( W \). The system is built so that there is no population leakage: the sum of each columns of \( W \) gives zero independently: \( \sum_{\nu' \neq \nu} w_{\nu' \to \nu} = w_{\nu} \), by definition. The eigenvalues satisfy \( \lambda_k \leq 0 \). In the long-time limit, all the exponential will converge to zero, except for the eigenvalue \( \lambda_1 = 0 \) (the position of the eigenvalue is arbitrary), where \( \exp(\lambda_1 t) = 1 \). The exponential matrix has the form \( \exp(\Lambda t) = \text{diag}(1, 0, \ldots, 0) \). The inverse transformation \( U^{-1} \cdot \text{diag}(1, 0, \ldots, 0) \cdot U \cdot N(0) \) is independent of \( t \) and gives directly the stationary solution.

This solution may be directly obtained by letting \( dN(t)/dt = 0 \) in the differential equation. This yields:

\[
W \cdot N = 0. \tag{6.18}
\]

The eigen-vector \( \mathbf{b}_1 \) associated with the zero eigen-value is proportional to the solution. But the later have a special norm. If \( N_s \) is the total number of electrons in the system, we have to ensure that \( \sum_{\nu} N_{\nu}(t) = N_s \). The usual norm of the eigen-vector is \( \| \mathbf{b}_1 \| = 1 \). For a real-valued vector: \( \sum_{\nu} b_{1,\nu}^2 = 1 \). The populations are therefore obtained by:

\[
N_{\nu}(t) = b_{1,\nu}^2 N_s \quad \text{where} \quad W \cdot \mathbf{b}_1 = 0 \quad \text{and} \quad \sum_{\nu} b_{1,\nu}^2 = 1. \tag{6.19}
\]

This solution may be adapted to describe the dynamics of the electrons in the whole structure by ensuring special boundary conditions (6.3). The latter essentially ensuring that electrons in the lowest subbands are pumped in the upper laser-state. This configuration models the coupling between adjacent periods.

It is worth to mention that diagonalisation of the scattering matrix solves the system without an external driving (inhomogeneous) term. As the problem consists of \( N \) linear equations with \( N \) unknown, the solution is trivial if we solve the homogeneous equations.

6.1.5 Inhomogeneous rate equations

Before we describe more complicated configurations, the simple rate equations can be coupled to an external source of carriers. This is modeled by pumping some subbands with a current density \( \mathbf{J} \). The net current being given by the sum of all individual currents \( \sum_{\nu} J_{\nu} \). We need to clarify the situation: if we allow a single subband to be pumped by a current density \( J \), the total population of the \( N \)-subband system will increase in time as there is no leakage to some external states. In particular the long-time limit diverges and there is no stationary solution. The amount of current injected in the system needs therefore be extracted from some other subbands: the current vector components need
therefore to satisfy $\sum_{\nu} J_{\nu} = 0$. The rate equations take the form:

$$\frac{dN(t)}{dt} = W \cdot N(t) + \frac{1}{q_0} J,$$

where $q_0$ is the signless elementary charge. If we look directly to the stationary solution, we impose $dN(t)/dt = 0$ and we have the equations: $q_0 W \cdot N = -J$. The system cannot be solved by a matrix inversion as $W$ has a zero eigen-value and therefore its inverse does not exist. But since the system consists of $N$ linear equations with $N$ unknown and an inhomogeneous term $-q_0^{-1} J$, a non-trivial solution exists and can be obtained by solving the $N$ linear equations. It is worth to remark that a solution does not exist necessarily in all systems. For example we may have situations in which the system is over-determined or ill-formed (isolated subbands). In such cases the stationary solution does not exist. Typically, if we solve $N(t)$ the populations of some states increase with time and the long-time limit is divergent.

The time evolution of the populations $N(t)$ is obtained by the integration of the differential equation for $N(t)$: this is done by varying the constants. The general solution is the sum of the general homogeneous solution and a particular solution of the inhomogeneous equation. The homogeneous general solution was already obtained: $U^{-1} \cdot \exp(\Lambda t) \cdot U \cdot A_0$, where $A_0$ is the integration constant vector at $t = 0$. We therefore transform $A_0 \rightarrow A_0(t)$ and we insert it in the inhomogeneous equation. This yields a differential equation for $A_0(t)$:

$$\frac{dA_0(t)}{dt} = \frac{1}{q_0} U^{-1} \cdot \exp(-\Lambda t) \cdot U \cdot J.$$  \hspace{1cm} (6.21)

The general solution to the inhomogenous problem is therefore known. We can integrate the above equation to yield:

$$N_{\text{inhom}}(t) = q_0^{-1} U^{-1} \cdot \exp(\Lambda t) \left( \int_0^\infty d\tau \exp(-\Lambda \tau) \right) \cdot U \cdot J$$

$$= q_0^{-1} U^{-1} \cdot \text{diag} \left( \frac{1}{\lambda_1}(e^{\lambda_1 t} - 1), \ldots, \frac{1}{\lambda_N}(e^{\lambda_N t} - 1) \right) \cdot U \cdot J.$$  \hspace{1cm} (6.22)

The eigen-values satisfy $\lambda_k \leq 0$. For all negative eigen-values, the elements of the diagonal matrix have a well-defined long-time limit: $-\lambda_k^{-1}$. But if we assume that the zero eigenvalue is $\lambda_1 = 0$, the long-time limit gives $t$ as matrix element. The total population in the system will increase with time and the stationary solution does not exist. Although in some subbands the population will converge to a stationary value, the sum does not: $\sum_{\nu} N_{\nu} \propto t$.

The general solution to the inhomogeneous equation reads:

$$N(t) = U^{-1} \cdot \exp(\Lambda t) \cdot U \cdot N(0) + \frac{1}{q_0} U^{-1} \cdot \text{diag} \left( \frac{1}{\lambda_1}(e^{\lambda_1 t} - 1), \ldots, \frac{1}{\lambda_N}(e^{\lambda_N t} - 1) \right) \cdot U \cdot J.$$  \hspace{1cm} (6.23)
This solution enables to examine the time evolution of the populations with a pump current \( J \). The convergence of the solution on a stationary solution is not explicit from the general solution and a deeper analysis of particular systems is required. Basically it depends on how the states are connected by scattering: we have to prevent closed loops or isolated states otherwise the model may yield unphysical results like zero, negative or infinite populations.

However if the states are all connected by scattering matrix elements and \( \sum_\nu J_\nu = 0 \), the populations \( N(t) \) will converge to a stationary solution.

### 6.1.6 Rate equations with hidden states

In the last two paragraphs (6.1.4-6.1.5) we have developed models for \( N \) subbands. The system is exact as there is \( N \) linear equations for \( N \) unknown and the sum-rule holds for each column \( \sum_\nu^N w_{\nu \rightarrow \nu'} = w_\nu \).

The latter is a very strong condition on the system. Such an exact modeling is efficient for developing general transport models where the populations in all states need to be known. The precise dynamics between the states needs therefore to be explicitly modeled with rate equations for each state and appropriate boundary conditions need to be imposed.

For simpler models it is somewhat useful to reduce modeling to a few states (upper and lower laser-states for example). In this case we need to keep an explicit dynamics for the interesting states and hide the rest. Nevertheless the scattering rates in the reduced set of states have to be computed with all states. The total scattering rate from a visible state \( \nu \) will have two contributions: \( w_\nu = w^{\text{visible}}_\nu + w^{\text{hidden}}_\nu \). The first term is the total rate from state \( \nu \) to all other visible states, while the second term is the total rate from the visible state \( \nu \) to all hidden states.

This reduction is accurate only if all the transitions from the hidden states to the visible states can be neglected with respect to the explicit scattering rates between visible states. If it is not the case, the reduction will yield a very bad approximation.

At the level of the scattering matrix \( W \) the previous reduction is obtained by a simple truncation of the matrix. If we assume that we have \( M \) hidden states (for simplicity, the last \( M \) states in the matrix \( W \)) the truncated matrix \( \tilde{W} \) has dimension \((N-M) \times (N-M)\). For each column, the sum-rule is broken: \( \sum_\nu^{(N-M)} w_{\nu \rightarrow \nu'} \neq w_\nu \).

This technique has many advantages; without hidden states it is necessary to specify the dynamics between all states. At this basic level, the knowledge of the dynamics in the whole structure cannot be modeled accurately. Typically it is better to select a few states...
(the upper and lower laser-states for example) and to compute the scattering rates between them and a few neighboring subbands (without getting the populations for the latter).

Here we need to add a pump current (inhomogeneous term) that models the current injection in some subbands. But we do not have to specify how the current is extracted from the system assuming this occurs in hidden states. Therefore $\sum_{\nu}^{(N-M)} J_{\nu} \neq 0$ in general. Although we can still model systems where the current, or some part of it, is extracted explicitly from visible states.

Since we have explicitly broken the sum-rule $\sum_{\nu}^{(N-M)} w_{\nu \rightarrow \nu'} \neq w_{\nu}$, we do not expect to have in general (some degenerated systems may have one) a zero eigen-value that gives a stationary state without an external source of carriers. This is normal as electrons are leaking to hidden states: the trivial solution $N = 0$ is therefore the only stationary state without an external current.

We can take the long-time limit $t \rightarrow \infty$ in Eq.(6.23). As we do not have a zero eigen-value, the matrix $\exp(\Lambda t) \rightarrow 0$ and the homogeneous solution vanishes as expected. Since $\Lambda^{-1}$ and $W^{-1}$ exist we can write the solution in the long-time limit as:

$$N(t \rightarrow \infty) = -\frac{1}{q_0} W^{-1} \cdot J$$

The latter coincide with the stationary solution obtained from the equation of motion in the case of a reduced system. It is worth to remark that the stationary solution does not depend on an initial value $N(0)$, since the homogeneous term vanishes in the long-time limit. The total populations in the visible states $N_{\text{visible}}(J)$ is a function of the pump current $J$.

For a quantum cascade structure we can therefore predict the population inversion between the laser states by knowing scattering rates between the active region states and the experimental current at which the real device is driven. The main draw-back is that the current cannot be predicted until we apply periodic boundary conditions on the system.

### 6.2 Simple rate equations in a quantum cascade laser

We apply the rate equations derived in (6.1) to a typical mid-infrared quantum cascade laser [9]. The latter has already been discussed from the band-structure viewpoint in (4.3). Here we only model the populations for a few relevant states of the active region. Although the quantum states in a planar heterostructure are subbands, we have shown that the simplest Boltzmann equations yield first-order differential equations for the total population of the subbands. The coupling constants being given by the lifetimes, usually
computed from \( k = 0 \) in the initial subband or at a characteristic thermal energy.

We do not impose periodic boundary conditions. It will be the subject of the next sections. The carrier injection from the left injector region is therefore modeled with a pump current density \( J \) in the upper laser-state.

The active region is shown in Fig. (6.1) and some characteristic times are outlined. We

Figure 6.1: The active region of a two-phonon structure is shown. The rate equations are modeled between state 4 (upper laser-level), 3 (lower laser-level), 2 and 1. The driving current \( J \) is injected in the upper laser-level. Some scattering times are depicted: \( \tau_{43}, \tau_{32}, \tau_{21} \).
first consider the following set of rate equations:

\[
\begin{align*}
\dot{N}_4(t) &= -\frac{N_4}{\tau_4} + \frac{J}{q_0} \quad \tau_4 = \frac{1}{\tau_{43}} + \frac{1}{\tau_{42}} + \frac{1}{\tau_{41}} \\
\dot{N}_3(t) &= -\frac{N_3}{\tau_3} + \frac{N_4}{\tau_{43}} \quad \tau_3 = \frac{1}{\tau_{32}} + \frac{1}{\tau_{31}} \\
\dot{N}_2(t) &= -\frac{N_2}{\tau_2} + \frac{N_3}{\tau_{32}} + \frac{N_4}{\tau_{42}} \quad \tau_2 = \frac{1}{\tau_{21}} \\
\dot{N}_1(t) &= -\frac{N_1}{\tau_{11}} + \frac{N_2}{\tau_{21}} + \frac{N_3}{\tau_{31}} + \frac{N_4}{\tau_{41}}
\end{align*}
\]

We have neglected any possible upward transitions. In this precise case the model is very simple as the coupling between the states is introduced one after the others. We can therefore restrict the model to the following set of states without loss of precision: \{4, 3\}, \{4, 3, 2\}. It will obviously not be the case if the scattering rate \(w_{2\rightarrow 3}\) cannot be neglected for example: model \{4, 3\} would have less precision than model \{4, 3, 2\}. If we look at the rate equation for state 1: the population will increase with time if do not give an escape time \(\tau_{11}^*\) to other (hidden) states.

These rate models do not take into account the coupling of the lower states 3, 2, 1 to the next injector region. The total scattering times \(\tau_3, \tau_2, \tau_{11}^*\) suffer from this approximation. In the present case the population of states 4 and 3 are expected to be modeled quite accurately. But the population of state 2 and 1 are not very reliable as scattering time \(\tau_{11}^*\) is practically unknown as the escape time to the injector region.

At this level of modeling we only want to have an approximation of the population inversion. The latter being known it is possible in principle to compute the gain between the laser-states. We can therefore estimate the threshold current density in a very simple manner. Furthermore we can model the coupling with the light and estimate the corresponding power and slope efficiency.

The stationary solution is obtained by setting \(dN(t)/dt = 0\). In this case the populations \(N_\nu\) can be expressed with simple algebra. The advantage of these simple rate equations is that the population inversion can be obtained algebraically:

\[
\Delta N \equiv N_4 - N_3 = \frac{J}{q_0} \tau_4 \left(1 - \frac{\tau_3}{\tau_{43}}\right)
\]

The condition for having a positive population inversion is: \(\tau_{43} > \tau_3\). This typical example shows the ease of the modeling with partial rate equations. The precision of the result can certainly be increased by adding the upward scattering time \(\tau_{34}\), although in this system
it is rather long (it is infinite compared to the typical time scale of downward transitions). Our purpose is to illustrate the simplest rate-equation models and not to obtain accurate algebraic solutions.

In the table Tab.(6.1) we have computed the scattering times for the active region by considering all possible transitions between states 4,3,2 and 1 with LO-phonon interaction computed form the subband-edge of the initial subband (zero kinetic energy). We have to mention that the matrix does not include the special scattering time \( \tau_1^* \), as the latter was added artificially in order to avoid divergences. We remark that the initial set of rate equations in Eq.(6.25) is a faithful representation of the dynamics of populations in the active region. In particular the state 1 has an infinite lifetime and its inclusion in the system will give rises to an infinite population in the long-time limit, as the coupling to the injector region is not included.

The scattering matrix \( W \) is therefore known. In the following computations, we keep the basis sorting \( \{4,3,2,1\} \).

We first compute the stationary state without an external current: \( J = 0 \). The eigenvector corresponding to the zero eigen-value is simply \( b_1 = (0,0,0,1) \). Meaning that all the initial populations end in the state 1 which as an infinite lifetime.

The following time dependent cases are illustrated in Fig.(6.2):

- We compute \( N(t) \) with \( J = 0 \) (isolated system) as shown in Fig.(6.2a). All the sheet carrier density of \( 2 \times 10^{11} \) cm\(^{-2} \) is in state 4 initially. All the populations are transferred to state 1, as expected from direct diagonalisation of the system.

- We add an external current \( J = (J,0,0,0) \) without extracting it from any state. The value of the current is \( J = 2.5 \) kA/cm\(^2\). The time evolution of the populations

<table>
<thead>
<tr>
<th>to/from</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.50</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1.33</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1.38</td>
<td>0.2</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1.92</td>
<td>0.66</td>
<td>0.26</td>
<td>( \infty )</td>
</tr>
</tbody>
</table>

Table 6.1: Scattering time computed for a two-phonon active region at alignment field of 48 kV/cm. The LO-phonon interaction (emission and absorption) was computed from the subband-edges (\( E=0 \)) for a temperature of 300 K. The energy of the optical-phonon is 32 meV, the Ga\(_{0.47}\)In\(_{0.53}\)As value for lattice-matched materials. The total scattering time is indicated in red on the diagonal of the table.
Figure 6.2: Time dependent simulations of rate equations are shown for different cases. 
(a) The homogeneous solution is shown with initial value of $N(0) = (2 \cdot 10^{11}, 0, 0, 0)$ cm$^{-2}$. 
(b) The general solution $N(t)$ is shown with an external current of 2.5 kA/cm$^2$ injected in 
state 4, the initial values are the same as in (a). (c) The general solution $N(t)$ is shown for 
the reduced system with states 4,3 and 2 and same conditions as in (b). All populations 
converge to stationary values.
is shown in Fig.(6.2b). The population in state 1 is asymptotically linear in time, while the populations of state 4,3 and 2 converge to stationary values.

- We reduce the system (6.1.6) to state 4,3 and 2. The populations converge to stationary values as show in Fig.(6.2c). The population inversion between state 4 and state 3 is $\approx 6.91 \cdot 10^9 \text{ cm}^{-2}$.

This simple system illustrates conceptually the type of solutions we may expect from rate equations when they are applied to quantum cascade structures without periodic boundary conditions and with constant scattering rates. In next sections appropriated boundary conditions are introduced and with a little effort several interesting physical features can be modeled.

### 6.3 Rate equations with boundary conditions

In this section we discuss rate equations models with special boundary conditions that ensure the periodicity of quantum cascade structure. These models were introduced by P. Harisson et al. [24,45–47]. The fundamental concept is to consider three adjacent periods of a quantum cascade structure. Rate equations are then applied to all states and a periodicity condition on the solution is imposed. The main interest of the model is its ability to predict of current density in the structure with simple rate equations.

#### 6.3.1 Periodicity of the spectrum in a quantum cascade structure

If we consider one effective band only, we know that the spectrum of a quantum cascade laser under an applied constant and uniform electric field $F$ is periodic in space and in energy. Formally if we consider an infinite structure (in the envelop function approximation with an energy-dependent effective mass), an energy-eigen subband wavefunction $\psi$ for $k_\perp = 0$ needs to satisfy:

$$\left(-\frac{\hbar^2}{2} \frac{\partial}{\partial z} m(z, E) \frac{\partial}{\partial z} + V(z) - q_0 F z\right) \psi(z) = E \psi(z) \quad (6.27)$$

If the period length is $L_p$, we know that the potential (the band-profile) $V(z)$ satisfy: $V(z + n L_p) = V(z)$, $n$ being an integer. If we let $V_0 = -q_0 F L_p$ being the potential drop for a translation $z \to z + L_p$, we can define the periodicity of the mass by:

$$m(z + n L_p, E) = m(z, E + n V_0) \quad (6.28)$$
It inherits the periodicity from the potential $V(z)$, shown in Fig.(6.3), by definition in Eq.(3.20). We define a translation operator $T(\lambda)$. The latter acts on wavefunctions as: $T(\lambda)\psi(z) = \psi(z + \lambda)$. It can be expressed with the (crystal) momentum $P_z$ as:

$$T(\lambda) = \exp \left( i\lambda \frac{P_z}{\hbar} \right) \equiv \exp \left( \lambda \frac{\partial}{\partial z} \right).$$  \hspace{1cm} (6.29)

It commutes with $P_z$ but not with operators $A(Z)$ depending on the position operator $Z$. We rewrite Eq.(6.27) in the vectorial form. All the quantities are expressed with operators $Z$ and $P_z$. The wavefunction is replaced by the ket $|\psi\rangle$. We apply a translation to this equation by a left-multiplication by $T(\lambda)$. If we insert the identity $T(\lambda) T(-\lambda) = I$, this yields:

$$\left( \frac{1}{2} P_z T(\lambda) \frac{1}{m(Z,E)} T(-\lambda) P_z + T(\lambda)(V(Z) - q_0 F Z) T(-\lambda) \right) T(\lambda) |\psi\rangle = E T(\lambda) |\psi\rangle.$$  \hspace{1cm} (6.30)

In the above equation the operators transform as: $T(\lambda) A(Z) T(-\lambda) = A(Z + \lambda)$. This equality holds since the operator is a function of $Z$. We write the translated ket $|\varphi\rangle \equiv T(\lambda) |\psi\rangle$ and we have:

$$\left( \frac{1}{2} P_z \frac{1}{m(Z + \lambda, E)} P_z + V(Z + \lambda) - q_0 F Z - q_0 F \lambda \right) |\varphi\rangle = E |\varphi\rangle.$$  \hspace{1cm} (6.31)

The interesting case is when $\lambda = n L_p$, as $m(Z + nL_p, E) = m(Z, E + nV_0)$ and $V(Z + n L_p) = V(Z)$. If we write $E' = E + nV_0$ this yields:

$$\left( \frac{1}{2} P_z \frac{1}{m(Z', E')} P_z + V(Z) - q_0 F Z \right) |\varphi\rangle = E' |\varphi\rangle.$$  \hspace{1cm} (6.32)

If we have a solution $\psi(z)$ with energy $E_0$, the translated wavefunctions $\psi(z + n L_p)$ are solutions with energies $E_0 - nq_0 F L_p$. The spectrum of the infinite structure is therefore periodic in position and in energy, as show in Fig.(6.3). We can therefore limit the calculation of the spectrum (the eigen-energies) to an energy band of width $q_0 F L_p$.

We restrict the discussion to bound-states. We assume that we have solved Eq.(6.27) and found a bound-state at energy $E_0$. As the spectrum is periodic in energy, all other eigen-states are found in the open interval $(E_0, E_0 + nq_0 F L_p)$. Formally we can write the fundamental spectrum $sp(H)$ as:

$$sp(H) = \{ E_m | H(E_m) \psi_m = E_m \psi_m \text{ with } E_m \in [E_0, E_0 + n F L_p] \} \quad \text{where}$$

$$H(E) = P_z \frac{1}{2m(Z, E)} P_z + V(Z) - q_0 F Z \quad \text{with} \quad \langle z | Z \rangle \in \mathbb{R}. \hspace{1cm} (6.33)$$
6.3. Rate equations with boundary conditions

Figure 6.3: The periodicity in space ($L_p$) and in energy ($q_0 F L_p$) are illustrated on one band for a quantum cascade structure. The translations need to be performed together in order to keep invariant an infinite structure.

It is clear that the spectrum is empty when $F \to 0$ as the periodicity in energy is lost.

The definition in Eq. (6.33) raises important physical questions. For the eigen-energies $H\psi = E\psi$ the definition of $sp(H)$ is unambiguous. Even if the band-width $q_0 F L_p$ is very narrow, we can search eigen-functions that are well-defined over the whole real axis. Although for increasing fields the bound-states become more localized, we have to assume that the wavefunction is coherent over an infinite distance. This would be right if the structure was perfectly periodic and if there was no scattering. But as soon as we account for the scattering Hamiltonian $H_{\text{scatt}}$, the coherence length of the wavefunctions is strongly reduced. We therefore expect that a model build on wavefunctions with an infinite coherence length will yield unphysical solutions in some circumstances. It will be the case, and these failures will guide us to a more physical model.
6.3.2 Computation of the wavefunctions

In the previous part we have shown formally that the width of the eigen-energy spectrum is $q_0 FL_p$. Even if we know that all bound-states can be obtained by solving $H\psi = E\psi$ on the whole real axis, this method cannot be practically implemented. We need to perform some approximations. We basically limit the computation of the bound-states on a few periods.

Conversely the problem can be solved on the interval $[0, L_p)$ with a propagation of the wavefunctions over an infinite energy-range. The wavefunction $\psi_0(x)$ associated with an eigen-energy $E_0$ can be reduced to an arbitrary portion of space of width $L_p$. We can define segments of the wavefunction with $\psi_{0\mu}(z) \equiv \psi_0(z + \mu L_p)$. An eigen-state is therefore represented by the infinite set: $\psi \sim \{\psi_{0\mu}(z)|z \in [0, L_p), \mu \text{ integer}\}$. Special continuity conditions need to be applied at each boundary $z = 0$ and $z = L_p$. This method is formally equivalent to the previous one. But here we need to limit the computation to a few energy-bandwidths $q_0 FL_p$ in a numerical implementation.

The spectrum for bound-states and their associated wavefunctions is finally obtained by using both method and strong assumptions on the spectrum of a quantum cascade structure.

For wavefunctions with an infinite extend, the definition of the spectrum in Eq.(6.33) is exact and contains all bound-states’ eigen-energies. For increasing field-strengths $F$, the wavefunctions are more and more localized. However for low fields, although all the energies are in the bandwidth $q_0 FL_p$, their wavefunctions are associated with remote periods. Even if the best way of modeling the transport for infinitely coherent wavefunctions would be to couple subbands between neighboring bandwidths, this is not numerically practical. We therefore need to perform approximations. For any field strength running between 0 and the alignment field $F_0$, we assume that, in a quantum cascade structure, if the average value of an eigen-wavefunction $\psi_0$ is $\langle Z \rangle_{\psi_0}$, then its standard-deviation $\sigma(\phi_0) = \langle Z - \langle Z \rangle_{\psi_0} \rangle_{\psi_0}$ satisfy: $\sigma(\psi_0) < 3L_p$.

This is to say that we can compute the band-structure of infinitely coherent wavefunctions by solving eigen-states on a potential region constituted by three adjacent periods $n-1$, $n$ and $n+1$. This is illustrated in Fig.(6.4) for two field-strengths in a two-phonon structure. To maximize the precision on the wavefunctions, we choose the wavefunctions localized in the central period $n$.

We need to develop some kind of algorithm that selects only these wavefunctions. We can use a criterion on energy (periodicity): if $E_0$ is the energy of a wavefunction localized in period $n$, then we can remove states with an energy $E_0 \pm q_0 FL_p + \varepsilon$, the energy $\varepsilon$ is a
Figure 6.4: The fundamental spectrum for a two-phonon quantum cascade structure [9] with infinitely coherent wavefunctions is shown in the approximation of three adjacent periods. Two field-strengths are shown. (a) For the alignment field. (b) For a low-field strength. The basis of wavefunctions is shown in red. The bandwidth is illustrated in shaded orange. In both cases we have aligned the bandwidth with the energy of the ground-state of the central period \( n \).
numerical parameter that takes into account the variation on the eigen-energies due to the finiteness of the potential region. But this criterion alone is not sufficiently efficient. We need to use a second criterion based on the centroid of the wavefunctions. These criterions allow to select the eigen-states shown in red in Fig.(6.4).

After having selected these states in the three-period potential, we obtain a basis $B_n = \{E_m^{(n)}, \psi_m^{(n)}(z)\}$ of wavefunctions with an infinite coherence length associated with period $n$ (up to the previous approximations). We can then rebuild the basis of the infinite structure by translating the basis $B_n$ in space $\psi_m^{(n+p)}(z) \equiv \psi_m^{(n)}(z - pL_p)$ and in energy $E_m^{(n+p)} \equiv E_m^{(n)} - p(q_0 F_0 L_p)$ for all integers $p$ and for all $m$ up to the dimension $N = \dim(B_n)$ of the basis.
6.3.3 Rate equations with boundary conditions

In the previous section we have obtained an approximated basis for an infinite quantum cascade structure with perfectly coherent wavefunctions. We model transport in this structure. We consider an arbitrary period with only nearest-neighboring periods. Modeling transport under this approximation is accurate as the standard deviation of the wavefunctions is less than three periods. Moreover, the scattering matrix elements being proportional to the overlap of the wavefunctions the limitation to three periods holds.

The concept of the transport model is to describe the net populations in subbands with rate equations. We use the definition of the central basis $B_n$ of the previous section.

We first consider all intra-period transitions which give a set of rate equations. We then consider the transitions from states of $B_n$ to states of $B_{n-1}$ and afterwards the transitions from states of $B_n$ to states of $B_{n+1}$. We therefore obtain three sets of rate equations. A final identification will impose periodic boundary conditions and will allow to define a current density.

With rate equations we model the total population in a subband $\nu$ of basis $B_n$ with a function $N^{(n)}_{\nu}(t)$. We define the scattering rate $W^{(n\rightarrow p)}_{\nu\rightarrow \mu}$ as being the scattering rate from state $\nu$ in basis $B_n$ to state $\mu$ in basis $B_p$. For now we do not make assumptions on how the scattering rates are computed. However we assume that we can factorize the total population and the scattering rate as in (6.1.3).

If we consider scattering between adjacent periods, the time variation of $N^{(m)}_{\nu}(t)$ is given by:

$$\frac{d}{dt} N^{(n)}_{\nu}(t) = \sum_{\mu \neq \nu} W^{(n\rightarrow n-1)}_{\mu\rightarrow \nu} N^{(n-1)}_{\nu} - N^{(n)}_{\nu} \sum_{\mu \neq \nu} W^{(n\rightarrow n-1)}_{\nu\rightarrow \mu} + \sum_{\mu} W^{(n+1\rightarrow n)}_{\mu\rightarrow \nu} N^{(n+1)}_{\mu} - N^{(n)}_{\nu} \sum_{\mu} W^{(n\rightarrow n+1)}_{\nu\rightarrow \mu} + \sum_{\mu} W^{(n-1\rightarrow n)}_{\mu\rightarrow \nu} N^{(n-1)}_{\mu} - N^{(n)}_{\nu} \sum_{\mu} W^{(n\rightarrow n-1)}_{\nu\rightarrow \mu}. \quad (6.34)$$

In the above equation, the first line models the intra-period scattering, the second line the scattering between period $n$ and $n + 1$ (scattering to the next period) and the third line scattering between period $n$ and $n - 1$ (scattering to the previous period).

It is important to remark that the scattering matrix\footnote{In this section, to avoid confusion, we use boldface type for the matrices and roman type for the matrix elements.} $W$ does not depend on the absolute position of the period. We have:

$$W^{(m+s\rightarrow m)} = W^{(m\rightarrow m-s)} \quad \text{for } m, s \text{ integers}. \quad (6.35)$$
We use this property to rewrite the scattering matrix between two periods only. Here we choose periods $n$ and $n + 1$:

$$ W^{(n-1 \rightarrow n)}_{\mu \rightarrow \nu} = W^{(n \rightarrow n+1)}_{\mu \rightarrow \nu} \quad \text{and} \quad W^{(n \rightarrow n-1)}_{\nu \rightarrow \mu} = W^{(n+1 \rightarrow n)}_{\nu \rightarrow \mu} \quad (6.36) $$

Using the above equations and re-arranging terms in Eq. (6.34) yields:

$$ \frac{d}{dt} N^{(n)}_{\nu}(t) = -N^{(n)}_{\nu} \left( \sum_{\mu \neq \nu} W^{(n \rightarrow \mu)}_{\nu \rightarrow \mu} + \sum_{\mu} W^{(n+1 \rightarrow \mu)}_{\nu \rightarrow \mu} + \sum_{\mu} W^{(n \rightarrow n+1)}_{\mu \rightarrow \nu} \right) + $$

$$ + \sum_{\mu \neq \nu} W^{(n \rightarrow n)}_{\mu \rightarrow \nu} N^{(n)}_{\mu} + \sum_{\mu} W^{(n+1 \rightarrow n)}_{\mu \rightarrow \nu} N^{(n+1)}_{\mu} + \sum_{\mu} W^{(n \rightarrow n+1)}_{\mu \rightarrow \nu} N^{(n-1)}_{\mu}. \quad (6.37) $$

We now impose that the populations $N^{(m)}$ are the same in any period of the structure. We have:

$$ N^{(m+s)} = N^{(m)} \quad \text{for} \ m, s \ \text{integers.} \quad (6.38) $$

This implements implicitly the periodicity of the structure by imposing the periodicity of the solution. We can therefore express the dynamics in terms of the populations in a single period as $N^{(n)} = N^{(n-1)} = N^{(n+1)} \equiv N$. We therefore drop the period index.

In Eq. (6.37) the terms $W^{(n \rightarrow n+1)}_{\nu \rightarrow \nu} N_{\nu}$ in the first line is compensated by the same term in the second line. As for the term $W^{(n+1 \rightarrow n)}_{\nu \rightarrow \nu} N_{\nu}$. This yields:

$$ \frac{d}{dt} N_{\nu}(t) = -N_{\nu} \sum_{\mu \neq \nu} \left( W^{(n \rightarrow \mu)}_{\nu \rightarrow \mu} + W^{(n \rightarrow n+1)}_{\nu \rightarrow \mu} + W^{(n+1 \rightarrow n)}_{\nu \rightarrow \mu} \right) + $$

$$ + \sum_{\mu \neq \nu} \left( W^{(n \rightarrow \nu)}_{\mu \rightarrow \nu} + W^{(n \rightarrow n+1)}_{\mu \rightarrow \nu} + W^{(n+1 \rightarrow n)}_{\mu \rightarrow \nu} \right) N_{\mu}. \quad (6.39) $$

We define a general scattering matrix:

$$ \mathbf{W} \equiv W^{(n \rightarrow n)} + W^{(n \rightarrow n+1)} + W^{(n+1 \rightarrow n)} \quad (6.40) $$

Its diagonal matrix elements are given by the total scattering rates from states $\nu$:

$$ W_{\nu \rightarrow \nu} \equiv - \sum_{\mu \neq \nu} W_{\nu \rightarrow \mu} \quad (6.41) $$

The Eq. (6.39) reads:

$$ \frac{d \mathbf{N}}{dt} = \mathbf{W} \cdot \mathbf{N}. \quad (6.42) $$

Except that we were able to close the equations for this scattering model, the very important added-value of the model is that it enables the definition of a current density at the interface between two periods. The current density given by a population $N$ is obtained as the net scattered populations between two adjacent periods:

$$ J[N] = q_0 \sum_{\nu, \mu} \left( W^{(n \rightarrow n+1)}_{\nu \rightarrow \mu} - W^{(n+1 \rightarrow n)}_{\nu \rightarrow \mu} \right) N_{\nu}. \quad (6.43) $$
6.3. Rate equations with boundary conditions

However it is shown in [63] that the current density in quantum cascade lasers arises from coherences between subbands $\rho_{ij}(k)$ with $i \neq j$. In the present model, the quantum ingredients are the scattering matrix elements and the eigen-energy subbands used for the calculation of the latter. As we have assumed that the wavefunctions have an infinite coherence length, the scattering between them is here treated as a perturbation. The latter will therefore not reduce the coherence length of the wavefunctions: the scattering hamiltonian is not taken into account in the propagation of the wavefunctions.

In this simplified model, the current seems to arise from direct scattering between fundamental states. The latter being the minimal set of eigen-states which is repeated infinitely across the structure. If we analyze the situation from a quantum mechanical point of view, we have here made the approximation that the coherences between the states, typically across the injection barrier, have an infinite lifetime. In (6.4) we will examine the problem of two states coupled by a barrier quantitatively and we will find that the current flowing through the structure can be modeled by an effective lifetime. Its value at the resonance is $\tau_{\text{eff}} = 2\tau_3$ where, $\tau_3$ is the total lifetime of the state on the right side of the barrier (for positive field-strength and standard mid-infrared structures). The current does not depend on quantum parameters, like the coupling energy between states across the barrier. It is like if the current arises only from incoherent scattering rates.

This approximation has major drawbacks as shown in the following paragraphs. The infinite lifetime of the coherences implies that we need to have a perfect resonance in order to drive a current in a quantum mechanical model. We therefore expect that the present model will exhibit instabilities near resonances.

How the three-period rate equation model is solved

The Eq.(6.42) can be solved by using the methods we have developed in (6.1.4), either by solving the time dependent problem in order to examine the time-dependence of the populations and the current. Either by solving directly the stationary problem with the eigen-vector associated with the $\lambda = 0$ eigen-value of operator $W$. The normalization of such vector to the sheet carrier density $N_s$ in a period of the structure is explained in (6.1.4). We can remark that the current is directly proportional to the sheet carrier density: as $N_\nu \propto N_s$, it implies that $J \propto N_s$.

There exist a linear domain between the current density and the sheet carrier density until the Hartree (self-consistent) potential sufficiently bent the potential and therefore alter the dynamics of the model: the eigen-energies are modified and consequently the scattering matrix elements $W_{\nu \rightarrow \mu}^{(n \rightarrow m)}$. This effect can only be observed if the rate equations
in Eq.(6.42) are embedded into a self-consistent computation as described in (4.5).

The scattering operator $W$ can be first roughly estimated by computing the scattering rates from the subband-edge of the initial subbands. However this approximation is very rough. It is better to compute the scattering rates at a given thermal energy $k_b T_e$ in the initial subband. The effective temperature $T_e$ can be estimated or computed with a detailed balance model for the kinetic energy [40], or even deducted from the LO-phonon bath temperature by an empirical linear law [92, 105]. We can also impose intra-subband Fermi-Dirac distributions for the subbands of the fundamental basis $B_n$ as discussed in (6.1.2). In this case the steady-state populations need to be computed iteratively.

If we consider initially a uniform partition of the sheet carrier density between subbands:

$$N_s^{(0)} = N_s / M$$

where $M$ is the dimension of the basis and the index (0) is the iteration number (It is often more efficient to start from a thermal distribution in the period as discussed in (4.5.5)). The scattering operator $W^{(0)} \equiv W[N^{(0)}]$ is then computed. We obtain the solution from the eigen-vector associated with the eigen-value equal to zero: $N^{(1)}$. We then update the scattering operator $W^{(1)}$ and so on until a convergence is achieved on the populations. This yields the solution $^2 N^{(\infty)}$.

This self-consistent routine on the populations and the scattering operator is then embedded into the Hartree self-consistent routine. At the end we should be able to converge on both the populations and the eigen-energies.

The final step is the computation of the current density $J$ for the stabilized populations and scattering operator. We are therefore able to predict the current-density as a function of the applied electric field. The predicted $J - V$ curve can be directly compared with experiments. In particular, the various approximation schemes discussed above (scattering from subband edge, thermal energy, Fermi-Dirac distributions and Hartree self-consistent potential) can be tested independently.

We will not discuss here the results obtained with these approximations. We will discuss it for the effective model examined in (7.1).

**Typical solutions of the three-period rate equations, instabilities**

In the present work, we do not have implemented the three-period rate equations model in all the previously discussed approximations. The scattering rates were computed from the subband-edges of the initial subbands for LO-phonon interaction only. The Hartree self-consistent potential was computed for a thermal distribution of carriers in the fundamental subbands. We use the notation $N^{(\infty)}$ since the numerical solution at a finite order (to the desired accuracy) is the closest to the exact solution. We assume that the latter is obtained for an infinite number of iterations.
period as explained in (4.5.5). We have used the temperature of the phonon-bath for the thermal distribution.

We first have applied the model to a diagonal injection design as shown in Fig.(6.5). In retrospect the diagonal structures are basically the most appropriated for this model as for most of the dynamic-range there is not instabilities caused by ill-defined resonances.

We have obtained interesting results as shown in Fig.(6.6). But already for this structure the field-current-density curves show instabilities in some specific field-regions: 74 kV/cm and ≥86 kV/cm. The instabilities are characterized by large current peaks which are not observed in experiments although the experimental curves displayed in the figure stop about the first instability region. There is no experimental signatures of current

![Figure 6.5: Three periods of a diagonal structure (S1548) are shown at a typical operation field-strength of 50 kV/cm. The fundamental set of infinitely coherent wavefunctions is shown in red. The lateral periods are shaded in grey in order to highlight the central period. The injection mechanism is based on direct scattering from the ground-state of the previous period, shown in dashed black, to the states of the next period. The shaded-pink barrier shows the dopants’ position.](image-url)
instabilities for fields $\geq 74$ kV/cm. The predicted instabilities are therefore artifacts of the model.

If we look at Fig.(6.6b) the instability in the current curve corresponds to an *anti-crossing between two wavefunctions in the infinitely coherent basis* as shown in the energy-diagram in Fig(6.6d). The extension of the anti-crossed wavefunctions is very large (about $2L_p$) as expected. Since these states are almost identical their lifetimes are comparable and they are approximatively equally populated. The problem of the infinite coherence length resurfaces here: when an electron is scattered into one of these states, it interacts *immediately* with states localized one period length away. This acts as a shortcut inside the model: the *transit time*, defined by the average time taken by an electron to be scattered one period length away, decreases abruptly which is translated by a large current peak. The latter has no physical significance as the coupling energy between the states in anti-crossing is less than 2 meV while the typical broadening energy (deduced from the broadening of the luminescence) is about 10 meV. The coupled states are therefore very fragile and scattering avoid their large spatial coherent extension.

One might think that the problem would be solved by removing these instabilities from the final curve. It is however not the case. The instability in Fig.(6.6b) is due to two weakly-coupled states and therefore the electric field domain corresponding to anti-crossed wavefunctions has a small extension: the artificial current-peak is very sharp. But if we now consider the second instability-region for fields higher than 86 kV/cm in Fig.(6.6c), the situation is pretty more complicated: at least three states are involved as shown in the energy-diagram in Fig.(6.6e). The field-region corresponding to the resulting instability is therefore quite larger. We can say that the current curve is undefined for fields larger than 72 kV/cm. In general the current curve is fragmented by possibly large unstable field-regions.

In the diagonal structure in Fig.(6.6a) the current is stable for interesting electric fields and the results are surprisingly close to the experiment for the rough approximations made in the evaluation of the scattering operator. It is worth to mention that the low-field regions (and *a fortiori* the current at zero field) are not accurately predicted. There is often a non-negligible current offset at zero bias.

However the apparent stability of the current for reasonable field-strengths has motivated us to apply the model to far-infrared structures. For the latter there are many resonances between the fundamental periods for about all field-strengths and we failed to obtain a current curve. Similar, but less severe instabilities should arise in mid-infrared quantum cascade structures based on resonant injection. However we have not tested the latter with this model. For examples of such simulations, the reader is referred to the literature [46,47].
Figure 6.6: (a) The simulation of a diagonal-designed structure (S1548) is shown for a large range of electric fields. The simulated curves are in solid-purple and the measurements in dashed-red. Increasing temperatures are considered from 100 K (lower current) to 300 K (highest current). The simulation in the region of 30-70 kV/cm is quite accurate. (b) The band-structure for the instability at 74 kV/cm is inspected. The latter corresponds to a weak anti-crossing between two states as shown in the anti-crossing diagram in (d). (c) The band-structure is shown for the instabilities for fields $\geq 86$ kV/cm. It shows that at least three states are involved in anti-crossing as shown in (e), which is a source of a more complicated instability than in (b).
In conclusion the essential problems raised by the infinite coherence length has forced us to stop the development of such models to the privilege of density matrix models that include sequential resonant tunneling at some barriers (typically at the injection barrier).

6.4 Modeling transport across a barrier

The conclusion previous section was that wavefunctions with a large coherence length yield current instabilities in the current-voltage curve. They are basically caused by weakly coupled states with a large spatial separation. In order to avoid these unphysical states we need in principle to account for scattering in the computation of the wavefunctions. Formally and conceptually this cannot be done at the level of wavefunctions. One has to formulate the problem in terms of one-particle propagators (Green’s functions) of the many-body problem. The scattering can be modeled efficiently in terms of perturbative expansions of the dressed propagator. The quantum modeling of heterostructures, like quantum cascade structures, has been developed by A .Wacker et al. [64, 109, 113] in the Keldysh/Kadanoff-Baym model [41]. Rather than computing wavefunctions and approximated scattering rates with the Fermi’s Golden rule, the correlation function ($G^<$, lesser Green function) is computed. The latter generalizes the density matrix and basically all one-particle properties, like the current density or the optical gain (through the complex conductivity) can be extracted from it. The fundamental advantage of such a general model is than the results doe no depend on particular choice of representation. This is the major drawback of the models we will develop in this work. In the effective transport model examined in (7.1) the coherent effects are investigated only at some coupling barriers, like the injection and in many structures the extraction barrier, as shown in Fig.(4.4). But in between we assume that the wavefunctions are coherent and we compute the populations with rate equations. These models cannot take into account quantum effects at any position in the structure [63].

6.4.1 Basic model: two coupled states across a barrier potential

Before we can develop formally a transport model that accounts for resonant tunneling at several barriers we have to discuss the current between resonant states across a barrier potential as illustrated in Fig.(6.7a). The problem was originally solved by Kazarinov and Suris in their seminal paper on light amplification in a superlattice [55]. We will do the formal derivation for subbands in (6.6) and discuss second-order contributions to the current and later to the optical gain (8.2.4). Here we consider simpler models
that involves only zero-dimensional states. The latter can model the edge of subbands in a heterostructure. However the in-plane energy dispersion is not included here. We therefore consider envelop function equations with $k_{\perp} = 0$.

In the tight-binding approximation developed in (4.4) we considered two states $|L\rangle$ and $|R\rangle$ coupled by a barrier potential. The latter will be typically the injection barrier in a quantum cascade structure. The hamiltonian of the system can be written as:

$$H = H_0 + W \quad \text{where} \quad H_0 = \begin{pmatrix} E_L & 0 \\ 0 & E_R \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} S_L & \hbar \Omega \\ \hbar \Omega & S_R \end{pmatrix}$$

The $H_0$ part is diagonal in the energies of the uncoupled system $E_L$ and $E_R$, while the $W$ part couples the two-states with a coupling energy $\hbar \Omega$ and shift energies by $S_L$ and $S_R$. The shift-energies are taken into account only for sake of completeness. We are not able to treat them in the effective transport model developed in (7.1), as we approximate the coupling between states across a barrier by treating each pair of states independently. We therefore set $S_L = 0 = S_R$.

This system can be easily diagonalised and the coupled energies $E_{\pm}$ are given by:

$$E_{\pm} = \bar{E} \pm \frac{1}{2}\sqrt{\hbar^2 \Delta^2 + 4\hbar^2 \Omega^2}.$$
Where we have defined the mean energy $\bar{E} = (E_L + E_R)/2$ and the detuning energy $\hbar \Delta = E_L - E_R$. We can plot the eigen-energies $E_{\pm}$ as a function of $\hbar \Delta$ as shown in Fig. (6.7b). This gives the usual anti-crossing diagram that we already have used in (6.3.3) when we have examined instabilities in the current-voltage curve.

In the basis obtained with an infinite coherence length the eigen-energy states would be $|+\rangle$ and $|-\rangle$ associated to eigen-values $E_{\pm}$. The $|\pm\rangle$ can be expressed by linear superpositions of $|L\rangle$ and $|R\rangle$, as given in [20]. The latter are perfect coherent superpositions of the localized states $|L, R\rangle$: they are of the form $|+\rangle = c_L |L\rangle + c_R |R\rangle$. To reduce the coherence length in the system, we may act on the coefficients $c_{L,R}$. These questions require a clarification of the physics we want to model.

We are not interested here to model the localization of the wavefunctions due to disordered potential. These effects are although relevant for quantum cascade structures. We want to model the loss of coherence of an electron in a many-body system. In the latter the electron can be scattered by phonons, roughness, impurities, other electrons, etc. The loss of coherence is typically expressed by the loss of correlation between two particles in the time evolution of the system, even if we are only interested in steady-state.

The complete quantum modeling of dynamical systems can be achieved with out-of-equilibrium Green’s functions. Here we treat the problem in the density-matrix formalism. The latter can be extended in order to model the loss of coherence with a single characteristic time as shown below.

We first consider the Rabi formula [20] for the two-state system. If we have initially ($t = 0$) prepared the particle in the state $|L\rangle$, the probability of founding it in the state $|R\rangle$ at time $t$ is given by:

$$P_R(t) = \frac{4\Omega^2}{4\Omega^2 + \Delta^2} \sin^2 \left(\sqrt{4\Omega^2 + \Delta^2} \frac{t}{2}\right)$$

(6.46)

The electron oscillates between state $|L\rangle$ and state $|R\rangle$ with a fixed time period. It means that the time $\tau_0$ taken to reach the maximum of $P_R(t)$ gives the time taken by the electron to tunnel across the barrier potential. This characteristic time is given by:

$$\tau_0 = \frac{\pi}{\sqrt{4\Omega^2 + \Delta^2}}.$$  

(6.47)

We can define a current density by defining the population $N(t) = N_s P_R(t)$. The associated current is defined by $J(t) = q_0 (dN(t)/dt)$. We have:

$$N(t) = N_s \left(\frac{\pi}{2}\right)^2 \tau_0^2 \Omega^2 \sin \left(\frac{\pi}{2} \frac{t}{\tau_0}\right), \quad J(t) = q_0 N_s \Omega^2 \sin \left(\frac{(\pi/\tau_0)t}{\pi/\tau_0}\right)$$

(6.48)

We have defined the current adhoc. When we will treat the problem in the density-matrix formalism, we will use the rigorous quantum mechanical definition.
6.4. Modeling transport across a barrier

The current flow across the barrier is periodic. The period is given by the tunneling time $\tau_0$. Here we cannot obtain a stationary current. This is normal since we do not have introduced a scattering time anywhere in the model. The charge is therefore conserved and the model is still a perfect coherent system: the current oscillation is not damped.

6.4.2 Current damping and the density-matrix formalism

By introducing conceptually a characteristic scattering time $\tau_s$ in the model, we can define different transport regimes with the ratio $\theta = \tau_0/\tau_s$:

- $\theta \ll 1$: The scattering time is very long compared to the tunneling time. The electron performs many oscillations. In this limit case we may use the $|\pm\rangle$ states and model transport by incoherent scattering.

- $\theta \gg 1$: In this opposite case, the scattering is very strong and the electron cannot tunnel through the barrier potential. In this case the transport is better modeled between localized states $|L, R\rangle$.

- $\theta \approx 1$: This intermediate case motivate the construction of a more subtle model that can treat coherent quantum effects with scattering. Typically the electrons tunnel across the barrier and are localized by scattering in the $|R\rangle$ state.

The question is how the characteristic scattering time $\tau_s$ can be introduced in the model and what is the quantity whose dynamical evolution is affected

Conceptually we already have remarked that the loss of coherence in the system should appear as a damping of the current oscillation. The long-time limit is unknown, but since electrons are scattered randomly with a characteristic time $\tau_s$, an increasing fraction of the electrons do not contribute anymore to the coherent oscillation.

The quantum mechanical definition of the current is given by the Heisenberg equations of motion [20]. For an observable $A(t)$ we have:

$$\frac{dA(t)}{dt} = -\frac{i}{\hbar} [H, A(t)] + \frac{\partial A(t)}{\partial t}. \quad (6.49)$$

We are interested in computing $I = e \langle \dot{Z}\rangle$ which is the quantum mechanical current for electrons, where $e = -q_0$ ($q_0$ being the signless elementary charge). We have:

$$I = -e N_s \frac{i}{\hbar} \langle [H, Z]\rangle. \quad (6.50)$$
Where the sheet carrier density $N_s$ was introduced. The latter multiplies the definition of the current if we keep a normalization to unity for the kets. The matrix of the operator $Z$ in the basis $|L,R\rangle$ is diagonal and reads:

$$Z = \text{diag}(z_L, z_R),$$

where $z_L, z_R$ is the averaged position of the wavefunctions $\psi_L(z) \equiv \langle z | L \rangle$ and $\psi_R(z) \equiv \langle z | R \rangle$. The commutator gives:

$$[H, Z] = \begin{pmatrix} 0 & \hbar \Omega d \\ -\hbar \Omega d & 0 \end{pmatrix} \quad \text{where} \quad d = z_R - z_L.$$  \hspace{1cm} (6.51)

Here it is worth to remark that we have neglected the non-diagonal parts $z_{LR}$ and $z_{RL}$ of the position operator. As we have assumed two isolated subbands only coupled by the energy $\hbar \Omega$, this approximation holds. If we compute the total commutator $[H, Z]$ without neglecting the $z^*_{RL} = z_{LR} = \int dz \psi^*_L(z)z\psi_R(z) \equiv z_0$ elements (where we have assumed real wavefunctions for bound-states) we have:

$$[H, Z] = \begin{pmatrix} 0 & \hbar \Omega d + z_0 \hbar \Delta \\ -\hbar \Omega d - z_0 \hbar \Delta & 0 \end{pmatrix}$$  \hspace{1cm} (6.52)

We therefore have two contributions: the first is given by tunneling $\hbar \Omega$ and the second by dipole $z_0$. In our system we have: $\hbar \Omega d \gg |z_0 \hbar \Delta|$, excepted for very large detunings, where the current is essentially zero. Another source of confusion in this calculation is that we cannot call $z_0$ a dipole as we know that the oscillator strength between conduction band states have a special definition (3.2.1). In anyway we neglect the $z_0 \hbar \Delta$ contribution to the current.

Now we evaluate the average value $\langle [H, Z] \rangle$ on a general time dependent solution $|\Psi(t)\rangle$ expressed in the basis of the localized states. We write $|\Psi(t)\rangle = a(t) |L\rangle + b(t) |R\rangle$. This yields the current density $J \equiv I/d$:

$$J(t) = eN_s i\Omega(ab^* - a^*b) = -eN_s \Im C.$$  \hspace{1cm} (6.53)

We have set $C = a^*b$. The current time evolution is therefore related to the product $a^*b$. If we solve the time dependent Schrödinger equation, the coefficients $a$ and $b$ satisfy the equations of motion:

$$i\hbar \dot{a} = E_L a + \hbar \Omega b$$
$$i\hbar \dot{b} = E_R b + \hbar \Omega a$$  \hspace{1cm} (6.54)

These equations require in general a diagonalisation, the eigen-vectors being $|\pm\rangle$. In the coupled basis the solutions are of type $\exp(iE_{\pm}/\hbar t)$. Back in the localized basis, this yields oscillations.
Instead of solving $a$ and $b$ independently, we try to find the equation of motion for $C = a^*b$ since the latter govern the current. We obtain:

$$\frac{d}{dt} C = i\Delta C + i\Omega\left( P_L(t) - P_R(t) \right). \quad (6.55)$$

In the above equation we have identified the probability of the system to be in state $|L\rangle$ as $a^*a \equiv |\langle L | \Psi(t) \rangle|^2 \equiv P_L(t)$ and in state $|R\rangle$ as $b^*b \equiv |\langle R | \Psi(t) \rangle|^2 \equiv P_R(t)$. The latter satisfy the equations of motion given by computing $d(a^*a)/dt$ and $d(b^*b)/dt$. We have a closed set of equations for $P_R, P_L, C$ and $C^*$ with the conservation of the norm $P_L + P_R = 1$.

We focus on the equations of motion for $C$. If we multiply the system by $N_s$, the term $P_L(t) - P_R(t)$ is simply the population difference between state $|L, R\rangle$. The current is therefore driven by the population difference.

If we solve this equation of motion, we will obtain undamped current oscillations. The simplest modification would be to add a damping term to the equation of motion for $C$. This can be modeled with a characteristic time $\tau_\perp$:

$$\frac{d}{dt} C = i\Delta C + i\Omega\left( P_L(t) - P_R(t) \right) + \frac{C}{\tau_\perp}. \quad (6.56)$$

This additional term will cause a damping of the current oscillations on a time scale given by $\tau_\perp$. In the closed two-state problem the long-time limit yields a zero current density.

We therefore have shown that acting on the current require to act on the coherences $\rho_{ab} = a^*b = C$ of the density matrix $\rho$. The simplest natural framework to model the loss of coherence in quantum processes is the density-matrix formalism.

### 6.4.3 The density matrix

In this section, we introduce briefly the density-matrix operator $\rho$. For a pure quantum state $|\Psi\rangle$ the latter is given by: $\rho = |\Psi\rangle \langle \Psi|$. The density matrix allows to model statistical mixtures (mixed states) but here we will consider only pure states. We have the following central properties:

$$\langle A \rangle = \text{Tr} (\rho A) \quad \text{Tr}(\rho) = 1. \quad (6.57)$$

where $A$ is an observable.

The diagonal elements of the density matrix are the occupation numbers (the populations) $\rho_{aa}$ which give the probability of finding the system in state $a$. The diagonal elements $\rho_{ab}$ are called the coherences of the density matrix. They model coherent effects (like the current) between quantum states as discussed in (6.4.2).
The equations of motion for the density-matrix operator are obtained either with the Heisenberg equations or directly with the Schrödinger equation: both derivations are equivalents. For a hamiltonian \( H \) we have:

\[
\frac{i\hbar}{\partial t} \rho = [H, \rho] \equiv \mathcal{L} \cdot \rho \tag{6.58}
\]

In the latter equation we have defined the *Liouville super-operator* \( \mathcal{L} \). It will be discussed in (7.1.2). In the next sections we use the standard equations of motion for the density matrix.

### 6.4.4 Two-state system treated with the density matrix

For a two-state system defined by \( H \) in Eq.(6.44), the equations of motion for the density matrix elements are given by:

\[
\begin{align*}
\dot{\rho}_{11} &= -i\Omega (\rho_{21} - \rho_{12}) \\
\dot{\rho}_{22} &= i\Omega (\rho_{21} - \rho_{12}) \\
\dot{\rho}_{12} &= -i\Omega (\rho_{22} - \rho_{11}) - i\Delta \rho_{12} \\
\dot{\rho}_{21} &= i\Omega (\rho_{22} - \rho_{11}) + i\Delta \rho_{21}
\end{align*}
\tag{6.59}
\]

We have replaced the index \( L \) by 1 and the index \( R \) by 2. The quantum mechanical current can be defined in the density-matrix formalism. With \( \langle \dot{Z} \rangle = -i\hbar^{-1}\langle [H, Z] \rangle \), we have:

\[
I = eN_s \langle \dot{Z} \rangle = eN_s i\hbar^{-1}\text{Tr}(\rho[H, Z]) = -e d N_s i\Omega (\rho_{21} - \rho_{12})
\tag{6.60}
\]

The current is therefore driven by the difference between coherences, the imaginary part of \( \dot{\rho}_{21} \) as: \( \rho_{21}^2 = \rho_{12} \). The function \( \rho_{21} - \rho_{12} \) satisfy the differential equation:

\[
\frac{d^2}{dt^2} (\rho_{21} - \rho_{12}) + (\Delta^2 + 4\Omega^2)(\rho_{21} - \rho_{12}) = 0. \tag{6.61}
\]

The current density therefore satisfy the same equation:

\[
\ddot{J} + \omega_0^2 J = 0. \tag{6.62}
\]

The current density between the two localized states is as expected an oscillation at the Rabi frequency \( \omega_0 = \sqrt{\Delta^2 + 4\Omega^2} \). We can remark that the populations are driven by the current as their time-derivatives are proportionals to \( \rho_{21} - \rho_{12} \).
If we now follow the idea developed in (6.4.2) of adding a decay term to the coherences, we transform the equations for $\rho_{12}$ and $\rho_{21}$ in:

$$\dot{\rho}_{12} = -i\Omega(\rho_{22} - \rho_{11}) - i\Delta \rho_{12} - \tau_\perp^{-1} \rho_{12}$$

$$\dot{\rho}_{21} = i\Omega(\rho_{22} - \rho_{11}) + i\Delta \rho_{21} - \tau_\perp^{-1} \rho_{21}$$

(6.63)

where $\tau_\perp$ is the characteristic time scale on which the coherences decay. The latter is called dephasing time. If we calculate the equation of motion for $J$ we find:

$$\ddot{J} + \tau_\perp^{-1} \dot{J} + \omega_0^2 J = 0.$$  

(6.64)

This is the differential equation of a damped oscillator with a fluid-friction coefficient $\tau_\perp^{-1}$. Three damping regimes can be identified. It is sufficient to insert the Ansatz $J = e^{\beta t}$. The regimes are identified by using the ratio between the tunneling time $\tau_0$ in Eq.(6.47) and the dephasing time $\tau_\perp$.

- $\frac{1}{4\pi^2} \left( \frac{\tau_0}{\tau_\perp} \right)^2 < 1$ The current is over-damped. The solutions are exponentials. The current decays to zero without oscillations. The dephasing time dominates the tunneling time.

- $\frac{1}{4\pi^2} \left( \frac{\tau_0}{\tau_\perp} \right)^2 = 1$ This is the critical solution, the system takes the longest time to decay to zero without oscillations.

- $\frac{1}{4\pi^2} \left( \frac{\tau_0}{\tau_\perp} \right)^2 > 1$ The current is under-damped, the current oscillates while its envelop decays to zero.

The inclusion of a dephasing time for the coherences allows to models the loss of coherence in the system due to scattering. The dephasing time will be found to consist of two contributions, a pure phase contribution that comes from loss of phase-relation between electrons due to scattering in a many-body system and a contribution due to incoherent scattering between quantum states. The latter is called population relaxation time.

The two-state system we are studying is isolated and therefore the current cannot reach a non-zero long-time (steady-state) limit except if we are giving special periodic boundary conditions. It is the case when we will model a superlattice in a simple two-state problem (6.4.5).

The two-state system can be coupled to the environment by pumping the left-localized state (index 1) with a fixed current density $J_0$. In order to conserve the charge, we need to extract the same amount of current from the right-localized state (index 2). As before we
can obtain a differential equation of the time dependent current density across the barrier. The equation has the structure of a damped oscillator driven by a current. If we assume that the populations are pumped with a constant current density \( J_0 \). This yields:

\[
\begin{align*}
\dot{\rho}_{11} &= -i\Omega(\rho_{21} - \rho_{12}) + q_0^{-1}J_0 \\
\dot{\rho}_{22} &= i\Omega(\rho_{21} - \rho_{12}) - q_0^{-1}J_0
\end{align*}
\]  

(6.65)

We obtain the following equation for the current density \( J(t) \) across the coupling barrier:

\[
\ddot{J} + \tau^{-1}_\perp \dot{J} + \omega_0^2 J = 4\Omega^2 J_0.
\]  

(6.66)

In the derivation of the above equation, we have set \( N_s = 1 \) when we have substituted back the current density definition \( J = q_0i\Omega(\rho_{21} - \rho_{12}) \). The multiplication factor \( N_s \) was introduced before since we assumed that \( \text{Tr}\rho = 1 \). By adding a pump current in the differential equations for the populations, we implicitly assumed that the populations \( \rho_{aa} \) of the density matrix give directly the electron concentration as the two-state problem does not conserve the norm anymore. We have therefore to set \( N_s = 1 \) in the definition of the current.

Some numerical results are shown in Fig. (6.8) for the under-damped regime and illustrate the discussion. It is worth to mention that with an external pump current, the normalization of the two-state system \( \text{Tr}\rho \) depends on \( J_0 \). The initial conditions need also to be chosen carefully to avoid unphysical results, like negatively populated states. The current curve \( J(t) \) can be obtained directly by solving Eq. (6.66) with initial conditions \( J(0) = 0 \) and \( \frac{d}{dt}J(0) = 0 \). We have chosen a zero detuning energy so that the system is able to sustain the pump current \( J_0 \). If we consider non-zero detuning, the amount of current driven by the two-state system is smaller than the pump-current and therefore carriers accumulate in the left-state and the right-state is completely depleted. In this simple model the validity of the results is guaranteed only on a finite time scale.

As we have already observed with rate equations, the boundary conditions are crucial. We will therefore examine them carefully in order to model a system where periodicity is well-implemented. It will be applied to periodic heterostructures in steady-state, as shown below.

### 6.4.5 Two-state system with periodic boundary conditions

In this section we develop the simplest model in the density-matrix formalism that implements periodic boundary conditions. We consider two 0-dimensional quantum states, labeled 1 and 2 as in the previous section. We assume that they are coupled by some
6.4. Modeling transport across a barrier

Figure 6.8: Numerical results are shown for a two-state system pumped with an external current $J_0$. The coupling energy is $\hbar \Omega = 2$ meV and the detuning energy is $\hbar \Delta = 0$ meV (resonance condition). The dephasing time has been set to $\tau_\perp = 0.6$ ps. The system is under-damped. (a) The current density $J(t)$ driven by the two-state system is shown until the steady-state value is reached at 500 A/cm$^2$, equal to the pump current. As expected the current has an overshoot to 800 A/cm$^2$ and decays with oscillations to the steady-state value. (b) The populations $N_{L,R}(t)$ in left/right states. The initial value of both was chosen to $N_0 = 8 \cdot 10^8$ cm$^{-2}$ and represents equally populated states in a thermal distribution, before the levels are pumped. The steady-state values give a population difference of $\Delta n \approx 2.82 \cdot 10^8$ cm$^2$.

barrier potential, with a coupling energy of $\hbar \Omega$ and a detuning energy of $\hbar \Delta$. In order to impose the periodicity on this system, we couple state 2 to state 1 by a scattering time $\tau_2$ and state 1 to state 2 by a scattering time $\tau_1$.

The superlattice illustrated in Fig. (6.9) can be modeled by this two-state system. We have
Figure 6.9: Three periods in a superlattice structure are shown. The band-structure was obtained for Ga_{0.47}In_{0.53}As /Al_{0.48}In_{0.52}As lattice matched on InP with wells of 8.0 nm and barriers of 0.25 nm under an electric field of 180 kV/cm. In each period two bound-states are confined. The ground-state \( 1_{n-1} \) is brought in resonance with state \( 2_n \). The coupling energy is \( \hbar \Omega \) and the detuning is \( \hbar \Delta \). Intra-period transitions are allowed by scattering times \( \tau_1 \) and \( \tau_2 \).

seen that rate equations with periodic boundary conditions can be written as in Eq.(6.42) provided that the scattering matrices account for intra-period scattering and scattering to nearest-neighboring periods as in Eq.(6.40). With two states we can therefore write \( W \) as:

\[
(W_{\nu \rightarrow \mu}) = \begin{pmatrix} -\tau_1^{-1} & \tau_2^{-1} \\ \tau_1^{-1} & -\tau_2^{-1} \end{pmatrix}.
\]

(6.67)

We can apply rate equations to the populations \( \rho_{11} \) and \( \rho_{22} \) of the density matrix. We consider the variations \( \delta \rho_{11} \) and \( \delta \rho_{22} \) caused by direct scattering:

\[
\begin{pmatrix} \delta \rho_{11} \\ \delta \rho_{22} \end{pmatrix} = (W_{\nu \rightarrow \mu}) \cdot \begin{pmatrix} \rho_{11} \\ \rho_{22} \end{pmatrix}
\]

(6.68)

Rate equations can therefore be included in a density matrix model with the identification \( \rho_{aa} = N_a(t) \). We can write the equations of motion for the density matrix \( \rho \). The latter is
6.4. Modeling transport across a barrier

the simplest improvement of rate equations between state 1 and state 2.

\[
\begin{align*}
\dot{\rho}_{11} & = -i\Omega(\rho_{21} - \rho_{12}) - \frac{\rho_{11}}{\tau_1} + \frac{\rho_{22}}{\tau_2} \\
\dot{\rho}_{22} & = i\Omega(\rho_{21} - \rho_{12}) + \frac{\rho_{11}}{\tau_1} - \frac{\rho_{22}}{\tau_2} \\
\dot{\rho}_{12} & = -i\Omega(\rho_{22} - \rho_{11}) - i\Delta \rho_{12} - \frac{\rho_{12}}{\tau_1} \\
\dot{\rho}_{21} & = i\Omega(\rho_{22} - \rho_{11}) + i\Delta \rho_{21} - \frac{\rho_{21}}{\tau_1}
\end{align*}
\]  

(6.69)

We have assumed a dephasing time \(\tau_\perp\). We have imposed periodic boundary conditions on states 1 and 2 by connecting states by incoherent scattering. We therefore have to impose the conservation of the norm: \(\text{Tr}\rho = 1\).

In the above equations, the detuning energy is \(\hbar\Delta = E_L - E_R\) as the density matrix equations were derived from the hamiltonian in Eq.(6.44). It corresponds to the detuning energy between states \(1_{n-1}, 2_n\) and states \(1_n, 2_{n+1}\). It is not the detuning between intra-period states \(2_n\) and \(1_n\). The transport across barriers are modeled by resonant tunneling with the coupling energy \(\hbar\Omega\) in \(H\), while the intra-period dynamics is modeled by rate equations. For computing the scattering times \(\tau_1\) and \(\tau_2\) we have to use states \(2_n\) and \(1_n\) and not states \(1_{n-1}\) and \(2_n\). There is no incoherent scattering between periods in this model. However incoherent scattering across the barrier would simply renormalize \(\tau_1\) and \(\tau_2\) as there are only two states in the model. The structure of the equations would not be modified.

We obtain the function \(\rho(t)\) by solving the density matrix equations in Eq.(6.69). Although the transition from initial conditions to a steady-state solution may be interesting, we want to obtain the steady-state solution directly. In these simple systems, the stationary solution \(\dot{\rho} = 0\) usually coincide with the steady-state solution. We can therefore solve the system algebraically.

From the equations for coherences we have: \(\rho_{21} + \rho_{12} = i\Delta\tau_\perp (\rho_{21} - \rho_{12})\). The latter can be substituted in the equation for \(\frac{d}{dt}(\rho_{21} - \rho_{12})\). It yields:

\[
0 = 2i\Omega\tau_\perp (\rho_{22} - \rho_{11}) - \left(1 + \Delta^2\tau_\perp^2\right) (\rho_{21} - \rho_{12})
\]  

(6.70)

The equations for \(\frac{d}{dt}\rho_{11}\) and \(\frac{d}{dt}\rho_{22}\) are redundant. We have to use one of them in order to express either \(\rho_{11}\) or \(\rho_{22}\), and use the conservation of the norm \(\rho_{11} + \rho_{22} = 1\). We obtain:

\[
\rho_{11} = \frac{1 + i\Omega\tau_2(\rho_{21} - \rho_{12})}{\tau_1 + \tau_2} \quad \text{and} \quad \rho_{22} = \frac{1 - i\Omega\tau_1(\rho_{21} - \rho_{12})}{\tau_1 + \tau_2}
\]  

(6.71)

We can therefore express \(\rho_{22} - \rho_{11}\) and substitute it back in Eq.(6.70). The current density \(J = q_0 N_s i\Omega(\rho_{21} - \rho_{12})\) reads:

\[
J = q_0 N_s \frac{2\Omega^2\tau_\perp}{1 + \Delta^2\tau_\perp^2 + 4\Omega^2\tau_\perp\tau_\perp} \left(\frac{\tau_1 - \tau_2}{\tau_1 + \tau_2}\right),
\]  

(6.72)
where we have introduced the total lifetime $\tau$ defined by:

$$\tau = \frac{1}{\tau_1^{-1} + \tau_2^{-1}} = \frac{\tau_1 \tau_2}{\tau_1 + \tau_2}. \quad (6.73)$$

The factor $(\tau_1 - \tau_2)/(\tau_1 + \tau_2)$ is constant and multiply the current resonance.

The current density $J(\hbar\Delta)$ as a function of the detuning energy $\hbar\Delta$ is a Lorentzian resonance centered at $\hbar\Delta = 0$. Its symmetry $J(-\hbar\Delta) = J(\hbar\Delta)$ allows to define a FWHM (Full Width At Half Maximum) denoted by $\Gamma_J$ as:

$$\Gamma_J = 2\hbar \tau_\perp \sqrt{1 + 4\Omega^2 \tau_\perp^2}. \quad (6.74)$$

The broadening of the current resonance is therefore a function of the dephasing time $\tau_\perp$, of the coupling energy $\Omega$ and of the total lifetime $\tau$. It is interesting to remark that the broadening of the current resonance depends of the total lifetime. Incoherent scattering between subbands will broaden the current resonance, regardless of the details of the dynamics between the states.

The constant factor $(\tau_1 - \tau_2)/(\tau_1 + \tau_2)$ is always positive as the state 1 is below the state 2 and therefore $\tau_1 > \tau_2$. This factor does not give the sign of the current. We assume that the superlattice is biased with the potential $-q_0 F z$. For $F > 0$ we expect a positive current, for $F < 0$ we expect a negative current. This is the case although it is not explicit from Eq.(6.72). However, the sign of the distance between the centroids of the wavefunctions $d = z_R - z_L$ changes: for $F > 0$ it is found positive, while for $F < 0$ it is found negative.

It is interesting to consider the current density as an effective transport time $\tau_{\text{eff}}$ across one period of the structure. We assume that the current density $J$ is positive. This enables to define $\tau_{\text{eff}}$ as:

$$J = \frac{q_0 N_s}{\tau_{\text{eff}}} \quad \text{with} \quad \tau_{\text{eff}} = f^{-1} \left( \frac{1 + \Delta^2 \tau_\perp^2}{2\Omega^2 \tau_\perp^2} + 2\tau \right) \quad (6.75)$$

where the constant factor was written defined as: $f = (\tau_1 - \tau_2)/(\tau_1 + \tau_2)$.

In the expression for $\tau_{\text{eff}}$, the first term is the time taken by an electron to cross the coupling barrier, this gives the tunneling time: $(1 + \Delta^2 \tau_\perp^2)/2\Omega^2 \tau_\perp$. It is not equal to the tunneling time in the Rabi formula in Eq.(6.47) as in the latter the dephasing time was not taken into account. However for $\Delta = 0$ expressions are closer. The second term $2\tau$ represents the scattering time inside the period. The factor 2 comes from the resonant tunneling process. Intuitively as the electrons oscillate between the states in resonance, only one-half in average can be scattered in regions on each sides of the barrier.

It is worth to remark that resonant tunneling effects at coupling barriers inside a periodic structure add at least (for $\Delta = 0$) a contribution $1/2\Omega^2 \tau_\perp$ to the effective time. The
transport time cannot be smaller than the tunneling time. This property is very important as large current instabilities arising in models based on incoherent scattering only cannot occur.

The condition for a positive population inversion is \( \rho_{22} > \rho_{11} \). This implies: \( \tau_2 > \tau_1 \). This condition cannot be fulfilled and this superlattice does not show a positive population inversion for any current densities \( J \).

The transport modeling in a superlattice based on this simple density matrix model works as a lorentzian resonance is predicted in the current curve when states \( 1_n \) and \( 2_n \) are brought in resonance. We will not go beyond this discussion for the superlattices and we apply the results to quantum cascade structures.

### 6.4.6 Resonant current formula and quantum cascade structures

The expression in Eq.(6.72) generalize the expression obtained in [16,88] where the authors have adapted the result of Kazarinov and Suris [55] to quantum cascade lasers. The latter is however more general as it accounts for the in-plane dynamics of subbands. This result will be furthermore generalized in (6.6).

The current density for a two-state system Eq.(6.72) can be used to describe a quantum cascade structure designed with a resonant injection. The state \( 1_n \) is identified with the injector ground-state and the state \( 2_n \) with the upper laser-level. Both are coupled by the injection barrier. It is usually assumed that \( \tau_1 \to \infty \) but it is not mandatory. However since \( \tau_1 \) connects by incoherent scattering the injector ground-state and the upper laser-level in the same period, it is clear that \( \tau_1 \to \infty \) for standard quantum cascade structures. It does not means that the injector ground-state has an infinite total lifetime. But in a two-state model the lifetime \( \tau_1 \) describe precisely the population transfer between state 1 and state 2. The scattering time \( \tau_2 \) is the total lifetime of the upper laser-level. The one-time formula is given by:

\[
J = q_0 N_s \frac{2\Omega^2 \tau_\perp}{1 + \Delta^2 \tau_\perp^2 + 4\Omega^2 \tau_\perp \tau_2}.
\] (6.76)

We can highlight different coupling regimes as discussed in [88] in the one-lifetime limit where \( \tau_1 \to \infty \) and therefore \( \tau \to \tau_2 \). For sake of completeness, we keep here the two-times formula. The original paper discusses the better parameters that optimize the operation of quantum cascade lasers. We perform this analysis, since on the one hand it is surprising that a simple two-state model enables relevant discussions for the design of the injection barrier, and on the other-hand, we want to examine the dependence of the broadening of
the current resonance in each regimes. We consider the maximal current $J(\Delta = 0)$. It reads:

$$J_{\text{max}} = J_{\text{max}}^{(0)} \left( \frac{\tau_1 - \tau_2}{\tau_1 + \tau_2} \right) \quad \text{with} \quad J_{\text{max}}^{(0)} = q_0 N_s \frac{2\Omega^2 \tau_\perp}{1 + 4\Omega^2 \tau_\perp \tau} \quad (6.77)$$

Without loss of generality we will discuss the $J_{\text{max}}^{(0)}$ term only.

(a) **Weak coupling regime**: In this regime, we assume that $4\Omega^2 \tau_\perp \tau \ll 1$. The order of magnitudes of $\tau_\perp$ and $\tau$ are the same, and they are shorter than $\Omega^{-1}$. It means that the tunneling time is dominated by scattering. The population transfer between $1_{n-1}$ and $2_n$ is therefore achieved by *incoherent* tunneling. This is manifest from the maximal current expression: $J_{\text{max}}^{(0)} \approx q_0 N_s 2\Omega^2 \tau_\perp$. The latter is driven by $\tau_\perp$ for a fixed coupling $\Omega$. In this special case the broadening of the current resonance is given by $\Gamma_J \approx 2\hbar \tau_\perp^{-\frac{1}{2}}$ and is basically given by the dephasing time.

(a) **Strong coupling regime**: This is the opposite regime where $4\Omega^2 \tau_\perp \tau \gg 1$. In this case, the maximal current is only controlled by the inter-subband scattering time $\tau$: $J_{\text{max}}^{(0)} \approx q_0 N_s / (2\tau)$. This is obtained with a high enough $\Omega$ and $\tau$ that compensates the short $\tau_\perp$. The dephasing time is basically a constant parameter for a pair of states in resonance across the coupling barrier. It cannot be easily tailored by design although it involves the scattering time $\tau$. The influence of the latter may not be negligible as it will be discussed in (7.1). In this approximation, the broadening of the current curve is given by: $\Gamma_J \approx 2\hbar \sqrt{4\Omega^2 \tau}$. The latter does not depend on $\tau_\perp$.

It is however worth to underline the limitations of this model. In the next paragraph we give a numerical application of the model to a reference two-phonon design. The Eq.(6.76) gives a largely overestimated current density at the injection resonance. It is important to remark that this widely-used formula gives fundamentally the current in a two-state system, as the normalization condition: $\rho_{11} + \rho_{22} = 1$, is always fulfilled. As shown in Eq.(6.70), the current $j$ (with $q_0 = N_s = 1$) between a pair of states is basically driven by the population difference $\rho_{22} - \rho_{11}$:

$$j = -\frac{2\Omega^2 \tau_\perp}{1 + \Delta^2 \tau_\perp^2} (\rho_{22} - \rho_{11}). \quad (6.78)$$

In a two-state system, the normalization enables to express the current with the standard formula derived previously. However if we now consider a typical mid-infrared quantum cascade structure is a transport regime, the electrons are distributed in the structure and the injection process does not involve all the sheet carrier density of the period. This reduces the magnitude of the injected current.
The formula in Eq.(6.76) is however accurate in weak-coupling regimes where the injected current is sufficiently low. The charge therefore resides essentially in the ground-state of the injector region. But for the strong-coupling regime we have to be very careful with this expression as it reduces the quantum cascade structure to a two-state system.

Application to the two-phonon design

We apply Eq.(6.76) to the two-phonon design in [9]. From (6.2) we know roughly the upper laser-level lifetime for a temperature of the LO-phonons bath of 300 K: \( \tau_2 \approx 0.5 \) ps. The coupling energy was computed between the ground-state of the injector and the upper laser-level in (4.4.4): \( \hbar \Omega \approx 2.53 \) meV. The dephasing time is chosen to a typical value of \( \tau_\perp \approx 0.2 \) ps. The latter being approximated by the FWHM (Full Width At Half Maximum) of luminescence measurements. The sheet carrier density per period is \( N_s = 2.06 \cdot 10^{11} \) cm\(^{-2}\) and the period length is \( L_p = 59.8 \) nm. The remaining parameter is the detuning energy \( \hbar \Delta \). The latter depends on the bias applied on the structure. It is therefore a function of the applied electric-field \( F_0 \). The detuning energy excepted, all other parameters are assumed to be constant with respect to the electric field. This is almost the case for the whole operation range of the laser.

In Fig.(6.10) we show the current curve as a function of the electric field, between 0 kV/cm and 100 kV/cm for various values of the dephasing time \( \tau_\perp \). The peak current is found near 28 kA/cm\(^2\) for \( \tau_\perp = 0.2 \) ps. This value is much too large compared to a typical value of 7 kA/cm\(^2\) in experiments. However we have assumed that all the charge is injected in the upper laser-state. It shows that it is a poor approximation and that a more precise transport model is required.

We check the coupling regime by computing: \( 4 \Omega^2 \tau_2 \tau_\perp \approx 30 \). The structure is therefore in a strong coupling regime and the maximal current is determined by \( \tau_2 \). The linewidth still have a dependence in \( \tau_\perp \) but the latter becomes less and less significant as \( \tau_2 \) increases.

6.4.7 Structure of the dephasing time \( \tau_\perp \)

As we have discussed in the previous sections the coupling of resonant states at a barrier requires a quantum mechanical modeling that goes beyond the simple rate equations. In the two-state density matrix model (6.4.5), the parameters were:

- The coupling energy \( \hbar \Omega \) and the detuning energy \( \hbar \Delta \).
- The dephasing time \( \tau_\perp \) that models the loss of coherence due to scattering.
Figure 6.10: The computation of the current density $J(F)$ as a function of the applied electric field $F$ is shown for a two-phonon design [9]. In inset, the detuning energy $\hbar \Delta(F)$ was computed as a function of the electric field. The resonance between the ground-state of the upper laser-level is found for $F_0 \approx 44.76 \text{kV/cm}$. The current density was evaluated for a range of dephasing times: 0.2 ps, 0.1 ps and 0.05 ps. The broadening of the resonant current curve is shown with shaded colors.

- The scattering times $\tau_1$ and $\tau_2$.

The coupling energy can be computed with a tight-binding model as in (4.4). The detuning energy is a free parameter. It can be connected to the applied electric field as shown in the inset of Fig.(6.10). The detuning energy was computed from the band-structure of the biased structure.

The scattering times $\tau_1$ and $\tau_2$ can be computed between all subbands for relevant interactions (5): LO-phonon, interface roughness, alloy disorder, ionized impurities, etc.

The dephasing time has a more complicated structure. We will show that it depends on the total lifetime $\tau$ of states involved in tunneling. However it has a pure phase contribution that requires independent calculations. If we consider the equations of motion for the coherences $\rho_{ab}$ and the populations $\rho_{aa}$, we have:

$$
\dot{\rho}_{ab} = -i\hbar[H, \rho]_{ab} - \hbar^{-1}\gamma_{ab}\rho_{ab} \quad \text{for } a \neq b
$$

$$
\dot{\rho}_{aa} = -i\hbar[H, \rho]_{aa} + W_{a \rightarrow a}\rho_{aa} + \sum_{b \neq a} W_{b \rightarrow a}.
$$

(6.79)
6.4. Modeling transport across a barrier

Where \( W_{a\rightarrow a} \equiv -\sum_{b\neq a} W_{a\rightarrow b} = -\Gamma_a / \hbar \): the total scattering rate from state \( a \) written in energy units. In the above equations for the coherences \( \rho_{ab} \) we have introduced the dephasing time (in energy units) \( \gamma_{ab} \).

We define \( C_n(t) \) as being the probability amplitude of finding the system in state \( a \). If we have an initial probability \( |C_a(0)|^2 \), then for increasing time (short-time limit) we will have:

\[
|C_a(t)|^2 \propto |C_a(0)| e^{-\Gamma_a t/\hbar}.
\]

(6.80)

This implies for the amplitude of probability:

\[
C_a(t) \propto \exp\left(-\frac{\Gamma_a}{2}\frac{t}{\hbar}\right).
\]

(6.81)

The dephasing time depends on the total lifetimes of the states involved in the coherence, here state \( a \) and state \( b \). The \( \frac{1}{2} \) factor comes from the above calculations. This factor is unimportant and is rather a matter of definitions. However the important fact is than the dephasing time depends on the total lifetime of the subbands given by inter-subband scattering and on a pure-phase contribution. In this work we have implemented the model developed by T. Ando [2] as already mentioned in (5.2.2).

As mentioned above, we have kept a pure phase contribution \( \gamma_{ab}^{(0)} \). The latter does not arise from a population relaxation but directly from dephasing between electrons performing the same intra-subband transition in different subbands, here subband \( a \) and subband \( b \). It is equivalent to the collision broadening in atomic systems [12]

The pure phase contribution can be computed from a particular hamiltonian \( H_{\text{scatt}} \). We have [2,3,103]:

\[
\gamma_{ab}^{(0)}(k) = 2\pi \sum \langle a| \mathbf{k} + \mathbf{q}| H_{\text{scatt}} | a\mathbf{k} \rangle - \langle b| \mathbf{k} + \mathbf{q}| H_{\text{scatt}} | b\mathbf{k} \rangle |^2 \delta (\varepsilon(\mathbf{k} + \mathbf{q}) - \varepsilon(\mathbf{k}))
\]

(6.82)

We have neglected the non-parabolicity of the conduction band by assuming a same effective mass for both states. The general expression is given in (5). The pure phase contribution \( \gamma_{ab}(k) \) is dependent on \( k \). It’s implementation in an effective 0-dimensional model will be discussed in (7.1).

If we sum all contributions from relevant interactions, we expect to obtain a good approximation for the pure phase contribution \( \gamma_{ab}^{(0)} \). However it seems that this approximation is not accurate for far-infrared lasers.

In these structures the electron-electron scattering may have an important impact.

However we have to mention that inhomogeneous broadening has to be taken into account for interface roughness (5.3.1). As the inhomogeneous broadening appears naturally on
optical linewidths it may also affect the current-voltage curve. We assume that in-plane islets of roughness contribute independently to the current density in far-infrared. In this case the effective current density results of the addition of parallel current densities.

As the roughness causes a variation of the well-widths, the resonance condition for a pair of subbands coupled by a barrier is reached at different electric field-strengths. This implies a broadening of the resonances in the current-voltage curves. This effect can partially explain the discrepancy between the predicted (too sharp) and the measured current-voltage curves for far-infrared structures.

Moreover some authors [14] prefer to fix the value of $\gamma_{ab}^{(0)}$ to some reasonable estimate extracted from measurements or from fitting techniques.

In the effective transport model presented in (7.1) we compute $\gamma_{ab}$ for each pair of subbands coupled by tunneling. In the next section we will examine a model for far-infrared structures where dephasing times are adjustable parameters.

### 6.5 A toy model for far-infrared single-quantum-well structure

As we have seen in (6.4.5) we may derive a tower of toy models in the density-matrix formalism. For example, we may consider various intermediate levels inside the period that may be adapted to the modeling of quantum cascade structures. Provided we implement periodic boundary conditions we will be able to derive a current density in terms of the internal parameters of the model.

An interesting model can be derived for quantum cascade designs where the injection and the extraction of carriers in/from the active region occurs by resonant tunneling. Single-quantum-well active regions can be accurately simulated with this model as demonstrated in [83] for a structure designed to emit in the far-infrared energy spectrum (15.5 meV). We have developed the simplest transport model that implements resonant injection and extraction of carriers from the active region. The light intensity was then simulated and coupled with the transport model. The resonant nature of the extraction of carriers yields physical features that was clearly identified in measurements. We were able to deduce the internal parameters (scattering and dephasing times) from the model with a high confidence. These scattering times were difficult to compute accurately with the models used at the time of the study.

In this section we will only consider the transport part of the model excluding the calcul-
lation of light. Although we may use numbers that were deduced from measurements of threshold current densities, we prefer to postpone the complete discussion in (8.3).
6.5.1 Calculation of the current density

The minimal model involves four states $g$, 2, 1 and $u$. The configuration of these states is sketched in Fig. (6.11). The pair of states $g$ and 2 models the injection from ground-state ($g$) to upper laser-level (2). We then consider scattering between states 2 and 1 (vertical transition in a single-quantum-well structure), 2 being the upper laser-level and 1 the lower laser-level. The extraction processes by resonant tunneling between state 1 and state $u$. The latter modeling the upper-state of a miniband in the injector region. We then close the equations by allowing incoherent scattering between state $u$ and state $g$.

As in (6.4.5) the coupling and detuning energies are computed for states adjacent to the coupling barrier, while the incoherent scattering times are computed between states inside the active and the injector regions. One period of the far-infrared structure examined in [83] is shown in Fig. (6.12).

We have a system with ten parameters. Each resonant doublet requires three parameters: a coupling energy $\hbar \Omega$, a detuning energy $\hbar \Delta$ and a dephasing time $\tau_\perp$. We have two scattering rates in the active region: $\tau_2$ and $\tau_1$ and two others in the injector region $\tau_u$ and $\tau_g$. Although these scattering times are the total scattering times for a given state, we set here $\tau_1, \tau_g \rightarrow \infty$, by assuming that they are negligible in comparison of $\tau_2$ and $\tau_u$. The number of parameters is therefore reduced to eight. The non-trivial density-matrix

![Figure 6.11](image.png)

Figure 6.11: The density-matrix model for resonant injection and extraction of carriers is sketched. The state are: $g$ the injector region ground-state, 2 the upper laser-level, 1 the lower laser-level and $u$ the upper state of the miniband in the injector region. The tight-binding parameters for the injection/extraction doublets are the coupling energies $\hbar \Omega_{i,e}$ and the detuning energies $\hbar \Delta_{i,e}$. The scattering times are $\tau_2$ for state 2 and $\tau_u$ for state $u$. The scattering times $\tau_1$ and $\tau_g$ are infinites. The model is closed by coupling state $u$ and state $g$ by incoherent scattering.
6.5. A toy model for far-infrared single-quantum-well structure

Figure 6.12: The structure studied in [83] is shown at the alignment field of $F = 3.12 \text{kV/cm}$. The Hartree self-consistent potential is evaluated for a thermal distribution of carriers in the period at a temperature of 50 K. The injector region is a superlattice while the active region in a single-quantum-well (SQW) where the optical transition occurs at 15.5 meV.

The equations are given by:

$$
\begin{align*}
\dot{\rho}_{gg} &= -i \Omega_i (\rho_{2g} - \rho_{g2}) + \frac{\rho_{uu}}{\tau_u} \\
\dot{\rho}_{22} &= i \Omega_i (\rho_{2g} - \rho_{g2}) - \frac{\rho_{22}}{\tau_2} \\
\dot{\rho}_{11} &= -i \Omega_e (\rho_{u1} - \rho_{1u}) + \frac{\rho_{22}}{\tau_2} \\
\dot{\rho}_{uu} &= i \Omega_e (\rho_{u1} - \rho_{1u}) - \frac{\rho_{uu}}{\tau_u} \\
\dot{\rho}_{g2} &= -i \Omega_i (\rho_{22} - \rho_{gg}) - i \Delta_i \rho_{g2} - \frac{\rho_{g2}}{\tau_{1,i}} \\
\dot{\rho}_{2g} &= i \Omega_i (\rho_{22} - \rho_{gg}) + i \Delta_i \rho_{2g} - \frac{\rho_{2g}}{\tau_{1,i}} \\
\dot{\rho}_{1u} &= -i \Omega_e (\rho_{uu} - \rho_{11}) - i \Delta_e \rho_{1u} - \frac{\rho_{1u}}{\tau_{1,e}} \\
\dot{\rho}_{u1} &= i \Omega_e (\rho_{uu} - \rho_{11}) + i \Delta_e \rho_{u1} - \frac{\rho_{u1}}{\tau_{1,e}} 
\end{align*}
$$

with $\rho_{gg} + \rho_{22} + \rho_{11} + \rho_{uu} = 1$. Where we have used the index $i$ and $e$ as a notation for the injection doublet $(g, 2)$ and
Chapter 6. Basic transport models and requirements

for the extraction doublet \((1, u)\). The equations are compatible with the conservation of the norm: 
\[ \rho_{gg} + \dot{\rho}_{22} + \dot{\rho}_{11} + \dot{\rho}_{uu} = 0. \]

We are interested in stationary solutions (the steady-state and stationary solutions coincide in this linear system) and we therefore set \(\dot{\rho} = 0\) and we solve the system algebraically. The equations for the coherences, for each doublet, yield:

\[
0 = -2i \tau_{\perp,i} \Omega_i (\rho_{22} - \rho_{gg}) + (1 + \Delta^2 \tau_{\perp,i}^2)(\rho_{2g} - \rho_{g2})
\]

\[
0 = -2i \tau_{\perp,e} \Omega_e (\rho_{uu} - \rho_{11}) + (1 + \Delta^2 \tau_{\perp,e}^2)(\rho_{u1} - \rho_{1u})
\]

The equations for \(\rho_{22}\) and \(\rho_{11}\) connect the resonant doublets with scattering time \(\tau_2\). This yields:

\[
\rho_{22} = i \Omega_i \tau_2 (\rho_{2g} - \rho_{g2}) \quad \text{and} \quad \rho_{22} = i \Omega_e \tau_2 (\rho_{u1} - \rho_{1u}).
\]

The polarizations are therefore equal:

\[
i \Omega_i (\rho_{2g} - \rho_{g2}) = i \Omega_e (\rho_{u1} - \rho_{1u}).
\]

It means that the current density driven by the doublet \((g, 2)\) is equal to the current driven by the doublet \((1, u)\) as expected from the charge conservation. The current density for both doublet is defined (for electrons \(e = -q_0\)) by:

\[
J_i = q_0 N_s i \Omega_i (\rho_{2g} - \rho_{g2}) \quad \text{and} \quad J_e = q_0 N_s i \Omega_e (\rho_{u1} - \rho_{1u})
\]

and we have \(J \equiv J_i = J_e\). The sheet carrier density \(N_s\) is factorized since we imposed \(\text{Tr} \rho = 1\).

We can express \(\rho_{gg}\) and \(\rho_{11}\) by using Eq.(6.85). We have:

\[
\rho_{gg} = \rho_{22} + T_i j \quad \rho_{11} = \rho_{uu} + T_e j
\]

Where we have introduced the tunneling time \(T_{i,e}\) and the unit current density \(j\) defined by:

\[
T_i = \frac{1 + \Delta^2 \tau_{\perp,i}^2}{2 \tau_{\perp,i} \Omega_i^2} \quad j = i \Omega_i (\rho_{2g} - \rho_{g2}) = i \Omega_e (\rho_{u1} - \rho_{1u}).
\]

With the equations \(\rho_{22} = \tau_2 j\) and \(\rho_{uu} = \tau_u j\), the normalization \(\text{Tr} \rho = 1\) gives:

\[
j \tau_{\text{eff}} = 1 \quad \text{where} \quad \tau_{\text{eff}} = T_i + 2 \tau_2 + T_e + 2 \tau_u.
\]

It means that the effective transport time \(\tau_{\text{eff}}\) is simply the sum of the injection time and the extraction time: each resonant path giving the value \(T + 2\tau\) as shown in Eq.(6.75). The current density is given by: \(J = q_0 N_s j\).
6.5.2 Analysis of the current

It is more intuitive to examine the effective transport time in Eq. (6.91) than the current density $J$. The transport time $\tau_{\text{eff}}$ is the sum of two resonances. It is worth to remark that they are not expected to occur at the same time. In the far-infrared structure mentioned before, the extraction resonance occurs (2.24 kV/cm) before the injector resonance (3.12 kV/cm).

In Fig. (6.13) the situation is illustrated for this far-infrared structure. The maximal current is not determined only by the injection resonance as in standard quantum cascade structures. As the injection and extraction doublets are not aligned at the same electric field, the minimum of $\tau_{\text{eff}}$ arises before the injection alignment. The maximal current is smaller than in the case where injection and extraction resonance conditions are met at the same field-strength. The parameters required for the computation of the figure were the scattering times inside the active region (SQW vertical transition) $\tau_2 = 12.2$ ps and

![Figure 6.13](image-url)

Figure 6.13: The effective transport time is computed for the far-infrared structure in [83]. In (a) the extraction resonance is reached at 2.24 kV/cm, the extraction tunneling time is the minimal. In (b) the injection resonance is reached at 3.12 kV/cm. The value $2(\tau_2 + \tau_u)$ is expected to be constant across the dynamic range. The minimum of $\tau_{\text{eff}}$ and therefore the maximal current is reached before the injection resonance.
inside the injector region $\tau_u = 8.9$ ps. The dephasing times are given at the injection barrier by $\tau_{\perp,i} = 1.6$ ps and at the extraction barrier by $\tau_{\perp,e} = 1.2$ ps. The coupling and detuning energies for each doublet were computed with the tight-binding model developed in (4.4) by varying the electric field $F$ as shown in Fig. (6.14). They were used to obtain the model parameters as a function of the applied electric field as the detuning energies $\Delta_{i,e}$ are required in the computation of $\tau_{\text{eff}}$.

Figure 6.14: The detuning and coupling energies for the injection/extraction doublets are shown. (a) The subband energies are plotted against the electric field $F$. In a single-quantum-well region, the transition energy between the first excited state and the ground-state is essentially constant with respect to $F$. The energies corresponding to the injector subband-edges are plotted and cross one after the others the lower laser-level 1 until the extraction resonance is met. The injector state $g'$ of the previous adjacent period is reported in order to illustrate the injection resonance. (b) The coupling energies are plotted against the electric field $F$ for the injection and extraction resonances. In the field-region where injection/extraction resonances occur, they are essentially constants.
It is interesting to remark that if the extraction and injection resonances are largely de-tuned and very sharp (long dephasing times $\tau_\perp$), $\tau_{\text{eff}}$ may have two minima. The region after the first one is therefore unstable (NDR) and the maximal current is given by the first minimum. The injection/extraction resonances are therefore decoupled from the transport viewpoint.

We can check in which regime are the injection and extraction resonances from the criterion of [88]. It is worth to remark that the strong coupling approximation requires that injection and extraction resonances occur at the same electric field, else the approximation is not very relevant. We have $\hbar \Omega_i \approx 0.8 \text{ meV}$ and $\hbar \Omega_i \approx 0.5 \text{ meV}$. We obtain:

$$4\Omega_i^2 \tau_{\perp,i} \tau_2 \approx 65 \quad \text{and} \quad 4\Omega_e^2 \tau_{\perp,e} \tau_u \approx 63. \quad (6.92)$$

If the injection and extraction resonances were matched the maximal current would be given by:

$$J_{\text{max}} = \frac{q_0 N_s \left( \tau_2 - \left( \tau_u + T_e \right) \right)}{2(\tau_2 + \tau_u)}. \quad (6.93)$$

If we examine Fig. (6.13), the minimal effective transport time is not far above $2(\tau_2 + \tau_4)$. For the maximal current the mismatched injection/extraction resonances have not a strong impact for the actual regime of parameters. However the situation is more complicated for the population inversion.

### 6.5.3 Population inversion

We examine the impact of the resonant injection/extraction on the population inversion in the active region (between state 2 and state 1). We compute $\Delta n/N_s = \rho_{22} - \rho_{11}$. By using Eq. (6.89) and $\rho_{22} = \tau_2 j$, we obtain:

$$\Delta n = \left( \tau_2 - (\tau_u + T_e) \right) \frac{J}{q_0}. \quad (6.94)$$

Even if the scattering times are constants, the population inversion is not just a constant multiplied by the current density $J$ as in (6.26). The lower laser-state lifetime due to incoherent scattering is infinite. The only way to build a positive population inversion is to extract carriers efficiently by resonant tunneling. This gives an effective lifetime for the lower laser-state: $\tau^* = \tau_u + T_e$. The problem is therefore reported on $\tau_u$ and the matching between the extraction and injection resonances.

In the case of the far-infrared structure under discussion, the extraction resonance occurs before the maximal current. The population inversion therefore shows a roll-over in stable electrical operation as shown in Fig. (6.15).
Chapter 6. Basic transport models and requirements

Figure 6.15: The roll-over in the population inversion $\Delta n$ is shown. The positive population inversion curve is outlined by a shaded area. The effective transport time $\tau_{\text{eff}}$ and the tunneling time $T_e$ for the extraction doublet are reported. The maximum of $\Delta n$ occurs before the NDR. The maximum of the population inversion does not coincide with the minimum of $T_e$. The population inversion is expressed as a fraction of the sheet carrier density $N_s = 3.84 \cdot 10^{10} \text{ cm}^{-2}$.

It is worth to remark that the maximum of $\Delta n$ does not correspond to the minimum of the tunneling time $T_e$ at the extraction barrier. Although the function $\tau_2 - (\tau_u + T_e)$ reaches its maximum when $\hbar \Delta_e = 0$, the maximum of $\Delta n$ is shifted by the multiplication by $J$.

This special feature is reflected in a laser structure by two threshold-currents. The first when laser action starts and the second (extinction threshold) when the laser actions stops due the roll-over in the population inversion. This feature have been observed in the light characteristic: We will discuss it later in (8.3.2).

6.6 Two subbands in the density-matrix formalism

In the past sections we always considered the subbands in heterostructures as being approximated by 0-dimensional quantum states. The in-plane dispersion being taken into account through effective scattering times as in (6.1.2). For transport models based on the density-matrix formalism the coherences $\rho_{ab}$ between 0-dimensional states were always
driven by the net population difference \( \rho_{bb} - \rho_{aa} \) between the subbands. From Eq.(6.70) the unity \((N_s = 1 \text{ and } q_0 = 1)\) current \( j = i\Omega(\rho_{ba} - \rho_{ab}) \) (the polarization) is driven by the net population difference:

\[
j = -\frac{2\Omega^2\tau_\perp}{1 + \Delta^2\tau_\perp^2} (\rho_{bb} - \rho_{aa}) .
\] (6.95)

This approximation is in fact pretty bad for modeling transport in quantum cascade lasers. If we consider the density resolved in \( k \)-space \( \rho_{ab}(k, k') \) (we lighten the notation by using \( k \) instead of \( k_\perp \)), the net population difference is given by:

\[
\rho_{bb}^0 - \rho_{aa}^0 = D_k^{(2)} \int d^2 k (\rho_{bb}(k) - \rho_{aa}(k)) ,
\] (6.96)

where the upper-index 0 indicated the 0-dimensional density matrix elements, and \( \rho_{bb}(k) \) are the density matrix elements diagonal in \( k \). The element \( \rho_{aa}(k) \) therefore gives the population in subband \( a \) at wave-vector \( k \). The constant \( D_k^{(2)} \) is the density of \( k \)-states in the reciprocal space, the index (2) indicating the dimensionality of the system. We will show here that \( k \) is not a good quantum number and as a consequence the current is not driven by the population difference at a given wave-vector but at a constant energy. This has been demonstrated in [108, 115] but neglected in [55]. For consistency with the literature and later sections, we name the current evaluated by the population differences at a constant wave-vector, first-order current, and the current computed at a constant energy, second-order current. This naming convention will become relevant in the derivation of both formula: in the second-order calculations we keep second-order terms that were previously averaged-out in the first-order calculations.

### 6.6.1 Failure of the first order current model

Before we obtain the second-order expression for the current we first show why it is required for an accurate modeling of the current.

**Ground-states in superlattices**

The simplest situation where the first-order current fails is in the calculation of the current driven between ground-states in a superlattice. We assume the latter to have thin enough wells so that only one energy-eigenstate can be found per period. We also assume thick enough barriers that make accurate the description of the transport by sequential resonant tunneling. We assume that a constant and uniform electric field is applied on the structure and that the latter is electrically stable. As there is no resonance between states across the coupling barrier, the requirement of electrical stability is basically fulfilled.
It is trivial on the experimental side that a current is driven in the superlattice in this configuration as there is no selection rules for any interactions that prevent transitions between ground-states of adjacent periods. If we assume a steady-state periodic solution, the ground-states are equally populated. The first order formula in Eq.(6.95) yields a zero current as: \( \rho_{bb} = \rho_{aa} \).

This simple example shows that the current between subbands cannot be modeled by a first order current.

**Leakage to excited states in quantum cascade structures**

Here the failure of the first order current model is more subtle. However it gives very bad predictions for the current and the population inversion in quantum cascade structures. We will not give a complete analysis of the problem. Our purpose is rather to examine a special case where the modeling of the current can alter significantly the population inversion. This occurs even if we are not considering equally populated subbands in contrast to the previous example. The situation illustrated in Fig.(6.16)

The quantum cascade structure is a mid-infrared single-quantum-well designed to emit at 7 \( \mu \)m. In this structure the carriers are injected by resonant tunneling in the upper laser-level and are extracted by resonant tunneling from the ground-state of the single quantum well active region. The situation is basically the same as in (6.5).

We focus on the coupling between the upper laser-subband (2) and a high-energy excited subband of the injector region (\( p \)). The coupling energy of 8.2 meV is quite large as the \( (p) \) state is confined in a few layers after the extraction barrier. As a comparison, the coupling energy between the ground-state of the active well (1) and the state \( u \) of the injector region (the extraction doublet) is only 2.7 meV. Regarding the detuning energy it is about \(-25.5\) meV between state (2) and \( (p) \) and essentially zero for the extraction doublet.

The coupling between the upper laser-subband and the state \( p \) opens a *parasitic channel* as it reduces the population inversion by extracting carriers from the upper laser-state. The modeling of the current it therefore crucial in the determination of the population inversion.

The unit current has the structure \( j = -\alpha(\rho_{bb} - \rho_{aa}) \) where \( \alpha \) may be regarded as the coupling strength between the current and the populations for a given resonant doublet. We name \( \alpha_e \) the coupling strength for the extraction from the ground-state and \( \alpha_p \) the extraction by the parasitic channel. At the alignment field we find: \( \alpha_p/\alpha_e \approx 0.154 \). It means that for equal population-differences in doublets, the leakage current represents
6.6. Two subbands in the density-matrix formalism

Figure 6.16: A period of a mid-infrared single-quantum-well active region (N257) is shown at the alignment field of $F_0 = 48 \text{ kV/cm}$. In this structure coherent tunneling occurs both at injection (4.2 nm) and extraction (3.0 nm) barriers. The active region (two quantum wells of 3.6 nm separated by a spike barrier of 0.2 nm) consists essentially in two bound-states. The doublet of the upper laser-subband (2) and a high-energy excited state of the injector region ($p$) has a coupling energy of 8.2 meV and a detuning energy of -25.5 meV. The doublet of the lower laser-subband (1) and the extraction state ($u$) of the injector region has a coupling energy of 2.7 meV and is in resonance.

13 % of the total current, which is a significant value.

To quantify the effect of the parasitic channel we model the system with rate equations for the total population of state 1,2, $p$ and $u$. In the equations we introduce a current $J_0$ that pumps the upper laser-state and we model the extraction from doublet with a first
order resonant current expression \( j = -\alpha (\rho_{bb} - \rho_{aa}) \). We have:

\[
\begin{align*}
\dot{N}_2 &= \frac{J_0}{q_0} - \frac{N_2}{\tau_{21}} + \alpha_p (N_p - N_2) \\
\dot{N}_1 &= \frac{N_2}{\tau_{21}} + \alpha_e (N_u - N_1) \\
\dot{N}_p &= -\frac{N_p}{\tau_p} - \alpha_p (N_p - N_2) \\
\dot{N}_u &= -\frac{N_u}{\tau_u} - \alpha_e (N_u - N_1)
\end{align*}
\]

(6.97)

Where \( \tau_{21} \) is the scattering time from the upper laser-state to the lower laser-state, we neglected upwards the transition \( 1 \rightarrow 2 \). The latter being negligible compared to all other time scales in the model. It is worth to remark that a positive current between state 2 and state \( p \) actually depletes state 2. This contribution should therefore be subtracted from the rate equation for state 2. The same convention is applied for the extraction doublet. The scattering rates \( \tau_u \) and \( \tau_p \) are the total scattering rates from state \( u \) and state \( p \) to all other states in the injector region.

We have computed the population inversion for the stationary solution \( \dot{N}_* = 0 \). It reads:

\[
\Delta N \equiv N_2 - N_1 = \frac{J_0}{q_0} \left( 1 - \frac{1 + \alpha_e \tau_u}{\alpha_e \tau_{21}} \right) \left( \frac{(1 + \alpha_p \tau_p) \tau_{21}}{1 + \alpha_p (\tau_{21} + \tau_p)} \right)
\]

(6.98)

An interesting function is the ratio \( R(x) \) between the population inversion computed with a fraction \( x \) of the parasitic coupling constant and the population inversion without the parasitic channel \( \alpha_p = 0 \). We have:

\[
R(x) \equiv \frac{\Delta N(\alpha_p \rightarrow x\alpha_p)}{\Delta N(\alpha_p = 0)} = \frac{1}{\tau_{21}} \frac{(1 + x\alpha_p \tau_p) \tau_{21}}{1 + x\alpha_p \tau_{21} + \tau_p}
\]

(6.99)

The ratio does not depend on \( \alpha_e \) and \( \tau_e \). The scattering times were computed for LO-phonon interaction (from subband-edges) at a temperature of 70 K. We have \( \tau_{21} \approx 1 \) ps, \( \tau_p \approx 0.82 \) ps and \( \tau_u \approx 0.41 \) ps. The coupling constants are given by: \( \alpha_p \approx 1.02 \cdot 10^{12} \text{ s}^{-1} \) and \( \alpha_e \approx 6.73 \cdot 10^{12} \text{ s}^{-1} \).

The ratio \( R(x) \) was plotted in Fig.(6.17) for \( x \in [0, 1] \). The effect of the parasitic channel has a clear impact on the population inversion. For \( x = 1 \), representing the first order current model, the population inversion is reduced by 36 %. The accuracy of the current modeling is therefore crucial. The ratio \( R(x) \) decreases as \( \tau_{21} \) increases, meaning that the population inversion is fragile with respect to the parasitic channels.
Figure 6.17: The population-inversion ratio $R(x)$ defined in Eq.(6.99) is plotted against the coupling strength $x$. The population inversion is reduced by the parasitic extraction channel. The parameters were computed for the single-quantum-well design (N257) for a temperature of the LO-phonon bath of 70 K.

We expect that the effect of the second-order terms in the current will be represented by an alteration of the coupling coefficient $\alpha_p$ of the parasitic channel. This means that the actual population inversion without assumption on the second-order current is somewhere between $x = 0$ and $x = x_0$, where $\alpha'_p = x_0 \alpha_p$ is the renormalized coupling coefficient. However we do not expect that the ratio $R$ will have the same analytic structure and that second-order current only act as a renormalization of the global coupling parameter $\alpha_p$. The results for second-order current modeling are given in (6.6.4).

For sake of completeness we give the condition for a positive population inversion: $\alpha_e(\tau_{21} - \tau_u) > 1$. For our parameters, the left-hand term is $\approx 3.9$.

### 6.6.2 Current between a pair of subbands

The simplest system where the current density can be modeled is a pair of subbands. The result presented here have been obtained independently by H. Willenberg [115] an A. Wacker [108]. Although the calculation techniques used by the latter are more efficient, we will perform the calculations in the density matrix formalism following the early work of Kazarinov and Suris [55].
The system consists of two subbands $|n\mathbf{k}\rangle$ with an energy dispersion given by:

$$\epsilon_n(\mathbf{k}) = \epsilon_n(0) + \frac{\hbar^2 k^2}{2m^*_n},$$  \hspace{1cm} (6.100)

where $n = 2, 1$. The dispersion relation was computed in the first-order approximation as in Eq.(3.25). The subbands are detuned by an energy of $\hbar \Delta = \epsilon_1(0) - \epsilon_2(0)$. We assume that they are coherently coupled with an energy of $\hbar \Omega$ for all $\mathbf{k}$. We drop any shift-energies.

We assume that subbands are orthogonal and normalized to unity:

$$\langle n\mathbf{k}|m\mathbf{q}\rangle = \delta_{n,m} \delta_{\mathbf{k},\mathbf{q}}$$  \hspace{1cm} (6.101)

If the subbands are conduction components in the effective two-band model (3.2) this relation does not holds. In the planar dimensions of the structure the orthogonality of the Bloch waves is preserved. However in the growth direction, the non-orthogonality of the conduction-band components will alter all matrix elements of operators, in particular the energy as discussed in (4.4).

If the parameters like the coupling energy are computed from the conduction-band envelop functions we assume that we have used the tight-binding method cited above. For dipoles we assume that we have used Eq.(3.2.1) with the special normalization given by the effective two-band model.

These parameters being computed, we can assume formally that the subbands are orthogonal and normalized to unity in a well-defined Hilbert space. The results of the model are therefore applicable the conduction-band components of heterostructure subbands. As we will basically consider transitions in the reciprocal space, the representation of operators is well-defined anyway.

Before considering explicitly only two subbands, we write the hamiltonian for a collection of $N$ subbands being coupled by energies $\hbar \Omega_{nm}$. It is worth to say that in this system the current density is more complicated to define than in the simples of the previous section. The $N$ subbands give a current density that depends of the position in the system. This the case in models studied by A. Wacker [63]. However the formalism used here is not appropriated for that. After the formal expansion for $N$ subbands, we will reduce the system to two subbands only and the current density will be defined between them. For treating more subbands, we have to know the spatial configuration of them, as in the

\footnote{If we go beyond the first order approximation, the $z$-dependent part of the envelop function depends on $\mathbf{k}_\perp$ in general, as well as the coupling energy.}
model for far-infrared single-quantum-well models (6.5), in order to define correctly the current operator.

We can split the hamiltonian into three parts: \( H = H_0 + H_\Omega + H_{\text{scatt}} \). The first yields the band-structure. It is basically the solution of the differential equation for the envelop function. We assume that:

\[
\langle n| k' | H_0 | m \rangle = \delta_{n,m} \delta_{k',k} \varepsilon_m(k) \quad (6.102)
\]

The second term \( H_\Omega \) contains the coupling energies between subbands. This contribution is diagonal in \( k \) but couples subbands \( n \) and \( m \) with an energy \( \hbar \Omega_{nm} \).

\[
\langle nk' | H_\Omega | mk \rangle = \delta_{k',k} \hbar \Omega_{nm}. \quad (6.103)
\]

If we assume that the subbands are spatially ordered, like ground-states in a superlattice, then we can consider only the coupling energies between nearest subbands \( \hbar \Omega \). The matrix elements reduce to: \( \langle nk' | H_\Omega | mk \rangle = \delta_{k',k} (\delta_{n,m-1} + \delta_{n,m+1}) \hbar \Omega \). This latter form will be used when we consider only two subbands.

The last term \( H_{\text{scatt}} \) models an elastic scattering potential. The later scatters the Bloch wave-vector \( k \) to \( k' \), where \( q \) is the exchanged wave-vector and may also cause direct scattering between subbands:

\[
\langle nk' | H_{\text{scatt}} | mk \rangle = \sum_q \delta_{k',k+q} V_{nm}^{q} \quad (6.104)
\]

The scattering potential \( V_{nm}^{q} \) is the Fourier transform of the potential \( V(z,r) \). If we write explicitly the braket, we have:

\[
\langle nk' | H_{\text{scatt}} | mk \rangle = S^{-1} \int d^2r e^{i(k' - k) \cdot r} \int dz \psi_n^* (z) V(z,r) \psi_m (z) = \sum_q \delta_{k',k+q} V_{nm}^{q} \quad (6.105)
\]

where \( S \) is the normalization of Bloch waves in the plane. If we assume that the scattering potential depends only on the in-plane position: \( V(z,r) = V(r) \), then the scattering hamiltonian is given by:

\[
\langle nk' | H_{\text{scatt}} | mk \rangle = \sum_q \delta_{k',k+q} V_q \quad (6.106)
\]

In this case we are considering a single intra-subband scattering potential for all subbands. The latter can model for example interface roughness. We are assuming that the latter are uncorrelated in the \( z \)-direction, as discussed in (5).
In the calculations we have to keep track of the tunneling and scattering coupling constants in order to stop the perturbation series at the second-order. For that it is necessary to distinguish the scattering potential $V^m_q$ in each subband. However, we do not consider inter-subband scattering. The scattering hamiltonian matrix elements read:

$$\langle n'k'|H_{\text{scatt}}|mk \rangle = \delta_{n,m} \sum_q \delta_{k',k+q} V^m_q$$ (6.107)

The matrix elements of the total hamiltonian $H$ therefore read:

$$\langle n'k'|H|m'k \rangle = \delta_{n,m} \delta_{k',k} \varepsilon_m(k) + \delta_{k',k} \hbar \Omega_{nm} + \delta_{n,m} \sum_q \delta_{k',k+q} V^m_q$$ (6.108)

We now consider the density matrix operator. In the matrix elements of the latter we can distinguish the diagonal part in $k$ written $\rho_{nm}(k)$. By considering these elements only, we are in the configuration of the previous 0-dimensional examples. The terms $\rho^{nm}(k)$ represents the population of state $n$ at wave-vector $k$, while $\rho^{mn}(k)$ are the coherences between state $n$ and $m$ at wave-vector $k$. If we neglect the scattering potential $V$, the system is diagonal in $k$. The solution for $\rho_{nm}(k)$ therefore coincide with the solution of $\tilde{\rho}_{nm}$ where the $k$ dependence is neglected. The current density is given by the first order formula by considering the integral in Eq.(6.96). This holds if we provide a dephasing time for the coherences. In 0-dimensional systems this time is introduced ad hoc. When we take into account the in-plane dynamics, the dephasing time will be related to the scattering potential as shown in the following calculations.

The scattering potential is therefore a key ingredient in the calculation. This is obvious for the superlattice example considered in (6.6.1): until scattering is not included in the system, the current is zero.

The density matrix elements $\rho_{nm}(k,k')$ for $k \neq k'$, represent intra-subband coherences between $k$ and $k'$ states for a given subband $n = m$. These elements for $n \neq m$ represent inter-subband coherences between state $|n\rangle$ and state $|m\rangle$. The tunneling energy $\hbar \Omega$ and the scattering potential $V$ will typically induce non-zero $\rho_{nm}(k,k')$ terms. The previous $k$-diagonal equations are therefore coupled between them. The current density computed from $\rho_{nm}(k)$ will therefore depends on $k$.

**Expression of the current density between a pair of subbands**

In the following calculation of the current, we neglect the non-parabolicity effects by setting $m_n \equiv m_w$ for all $n$. If we keep the non-parabolicity effects, the calculations are somewhat
unclear and cumbersome. We prefer to introduce the non-parabolicity effects on the final result. This method is much more simple and it yields an accurate solution.

Before we solve the density-matrix equations, we can assume to have found a solution \( \rho \) and we compute the corresponding current. From quantum mechanics we have seen that the current \( I \) is given (for electrons \( e = -q_0 \)) by:

\[
I = q_0 N_s \langle \dot{Z} \rangle
\]

\[
= -q_0 N_s \hbar^{-1} \text{Tr} (\rho[H, Z])
\]

This requires to define the position operator \( Z \). As we assume that the envelop function is the same for all \( k \), we have:

\[
\langle nk' | Z | mk \rangle = \delta_{nm} \delta_{k',k} z_n + \delta_{k',k} \delta_{nm}
\]

where \( z_n = \langle n | Z | n \rangle \) is the average position of the envelop function for subband \( n \). And \( z_{nm} \) the diagonal contribution. We set formally \( z_{nn} = 0 \) in order to ensure well-defined matrix elements. We know from Eq.(6.52) that the diagonal contribution to the current can be neglected and therefore we assume \( z_{nm} \rightarrow 0 \) without getting into cumbersome algebra. We then compute the matrix elements of the commutator \([H, Z] \):

\[
\langle nk | [H, Z] | mk \rangle = (z_m - z_n) \langle nk | H | mk \rangle.
\]

The trace \( \text{Tr}(\rho[H, Z]) \) gives:

\[
\text{Tr}(\rho[H, Z]) = \sum_{n,k} \langle nk | \rho[H, Z] | nk \rangle
\]

\[
= \sum_k \sum_{m,n} \hbar \Omega_{nm}(z_n - z_m) \rho_{nm}(k)
\]

\[
+ \sum_{k,q} \sum_{m,n} V_{qm}^{nm}(z_n - z_m) \rho_{nm}(k, k + q)
\]

The current seems to have two components: the first arises from the coherences and gives the usual resonant current. The second is due to direct scattering between subbands spatially separated by a distance \( z_n - z_m \). It is worth to remark that this term vanishes identically and there is no contribution to the quantum mechanical current from scattering [63]. This can be proven by computing the commutator of \( Z \) with the scattering potential \([H_{\text{scatt}}, Z] = [V(Z, R), Z] \equiv 0 \) as the scattering potential does not depend on the (crystal) momentum \( P_z \). Assumptions like \( V_{nm} = \delta_{n,m} V \) are of course not necessary in order to have a zero contribution of scattering to the current. However it is not obvious from the matrix elements in the last line of Eq.(6.112). As for the case of the current between ground-states in a superlattice (6.6.1), scattering is necessary for transport but does not drive the current.
The hamiltonian $H_0 + H_\Omega$ depends necessarily on the kinetic energy and it has therefore a non-trivial commutation relation with $Z$. We have:

$$I = -q_0 N_s i \hbar^{-1} \text{Tr} (\rho[H, Z])$$

$$= -q_0 N_s \sum_k \sum_{m,n} i \Omega_{nm} (z_n - z_m) \rho_{nm}(k)$$

$$= q_0 N_s \sum_{(m,n)} \sum_k i \Omega_{(mn)}(z_m - z_n) (\rho_{mn}(k) - \rho_{nm}(k))$$

$$= \sum_{(n,m)} \sum_k I_{nm}(k).$$

where the summation runs over pair of subbands $(m, n)$. We have assumed real coupling energies with $\hbar \Omega_{nm} = \hbar \Omega_{mn} \equiv \hbar \Omega_{(n,m)}$ and we have defined the $k$-dependent current $I_{nm}(k)$.

Now we restrict the problem to two subbands only: $\{ |1k\rangle, |2k\rangle \}$. The current density is defined by dividing $I$ by $z_2 - z_1$:

$$J = q_0 N_s \sum_k i \Omega(\rho_{21}(k) - \rho_{12}(k))$$

(6.114)

As discussed before, we assume that the scattering potential will modify the coherences as a function of $k$ and therefore the $k$-dependent current. This requires to the density-matrix equations.

**Steady-state solution with the Laplace average**

Before we give details of the calculations, it is worth to remark that in general a stationary solution, imposed by $\dot{\rho} = 0$, cannot be obtained. Even if the solution may be solved algebraically, as the Hilbert space is infinite, the convergence of the algebraic solution has to be discussed. It is more efficient to consider finite-size systems and try to diagonalise large sparse matrices modeling a particular choice of scattering potential. This method works in principle.

Here we will obtain an algebraic expression for the current by a different method.

The time-dependent problem cannot be solved algebraically. However there are ways in order to obtain directly the steady-state solution from the density matrix equations. An efficient technique uses the *Laplace average*. For a time depend function $f(t)$, the latter is defined by:

$$f(s) \equiv L[f](s) = s \int_0^\infty dt \ e^{-st} f(t).$$

(6.115)

The steady-state limit of $f$ is obtained by $s \to 0_+$. We can test this method on simple differential equations in order to convince ourselves that this method yields the expected results. This is discussed in more details in (C).
If we consider an ordinary differential operator $d/dt$, the Laplace-average transformation gives:

$$ L \left[ \frac{df}{dt} \right] = -sf(0) + sf(s) \tag{6.116} $$

The Laplace average therefore maps differential equations on algebraic equations with initial values: $f(0)$. The algebraic system is then solved for $f$. The limit $s \to 0_+$ is taken in order to obtain the steady-state solution.

The differential equation for the density matrix $i\hbar \dot{\rho} = [H, \rho]$, where $\frac{d}{dt}H = 0$, is mapped on an algebraic system:

$$ i\hbar \frac{d}{dt}\rho = [H, \rho] \to i\hbar sf(s) = i\hbar s\rho(0) + [H, f(s)]. \tag{6.117} $$

Here and in the following calculations we use the notation $f(s)$ for the Laplace-average transform of $\rho(t)$.

**Steady-state solution of the density matrix equations**

We first compute the matrix elements of $[H, \rho]$ in the basis $\{ |nk \rangle \}$. We have:

$$ \langle nk'| [H, \rho] | mk \rangle = (\varepsilon_n(k') - \varepsilon_m(k)) \rho_{nm}(k', k) + \sum_p (h\Omega_{np} \rho_{pm}(k', k) - \rho_{np}(k', k) h\Omega_{pm}) + \sum_q (V_{n-k}^* \rho_{nm}(q, k) - \rho_{nm}(k', q) V_{q-k}^*). \tag{6.118} $$

The equation of motion for $\rho_{nm}(k)$ read: $\rho_{nm}(k, k')$:

$$ i\hbar \frac{d}{dt}\rho_{nm}(k) = (\varepsilon_n(k) - \varepsilon_m(k)) \rho_{nm}(k) + \sum_p (h\Omega_{np} \rho_{pm}(k) - \rho_{np}(k) h\Omega_{pm}) + \sum_q (V_{k-q}^* \rho_{nm}(q, k) - \rho_{nm}(k, q) V_{q-k}^*). \tag{6.119} $$

$$ i\hbar \frac{d}{dt}\rho_{nm}(k', k) = (\varepsilon_n(k') - \varepsilon_m(k')) \rho_{nm}(k', k) + \sum_p (h\Omega_{np} \rho_{pm}(k', k) - \rho_{np}(k', k) h\Omega_{pm}) + \sum_q (V_{k'-q}^* \rho_{nm}(q, k) - \rho_{nm}(k', q) V_{q-k}^*). \tag{6.120} $$

In the last equation, we consider the Born approximation by replacing in the last line: $\rho_{nm}(q, k) \to \delta_{q,k}\rho_{nm}(k)$ and $\rho_{nm}(k', q) \to \delta_{q,k'}\rho_{nm}(k')$. This simplification closes the equations.

The $k$-diagonal density matrix $\rho_{nm}(k)$ has to be solved in order to compute the current density. We first need to obtain a solution for the non-diagonal part $\rho_{nm}(q, k)$ by solving
Eq. (6.120). In the latter we assume first order tunneling only by replacing: \( \rho_{pm}(k', k) \rightarrow \delta_{p,m} \rho_{mm}(k', k) \) and \( \rho_{np}(k', k) \rightarrow \delta_{p,n} \rho_{mm}(k', k) \) in the second line. We take the Laplace average on the whole equation and we assume vanishing initial coherences \( \rho_{nm}(k', k) \big|_{t=0} = 0 \). This yields:

\[
\frac{i \hbar s f_{nm}(k', k)}{\varepsilon_n(k') - \varepsilon_m(k)} = \left( \varepsilon_n(k') - \varepsilon_m(k) \right) f_{nm}(k', k) \\
+ \hbar \Omega_{nm} \left( f_{nm}(k', k) - f_{nn}(k', k) \right) \\
+ V_{k' - k}^{m} f_{nm}(k) - f_{nm}(k') V_{k' - k}^{m} \tag{6.121}
\]

In order to solve this equation, we first have to solve the intra-subband polarization \( f_{mm}(k', k) \). This yields:

\[
f_{mm}(k', k) = - \frac{V_{k' - k}^{m} \left( f_{nm}(k) - f_{mm}(k') \right)}{\varepsilon_m(k') - \varepsilon_m(k)} - i \hbar s \tag{6.122}
\]

In the steady-state limit \( s \rightarrow 0_+ \) we use the identity:

\[
\frac{1}{\omega - i s} \rightarrow \mathcal{P} \left( \frac{1}{\omega} \right) + i \pi \delta(\omega) \tag{6.123}
\]

Where \( \mathcal{P} \left( \frac{1}{\omega} \right) \) is the principal value of function \( 1/x \). In the sense of distributions, the above identity is defined on a smooth function \( \varphi(x) \) with a compact support:

\[
\lim_{s \rightarrow 0_+} \int_{-\infty}^{\infty} dx \frac{\varphi(x)}{x \pm is} = \mp i \pi \varphi(0) + \lim_{s \rightarrow 0_+} \int_{|x|>s} dx \frac{\varphi(x)}{x} \tag{6.124}
\]

In Eq. (6.122) the pole \( i \pi \delta(\omega) \) yields a zero numerator and we keep only the principal part by assuming that the numerator makes \( f_{mm}(k', k) \) converge to zero. We can therefore remove the \( \mathcal{P} \) to get:

\[
f_{mm}(k', k) = \frac{V_{k' - k}^{m} \left( f_{nm}(k) - f_{mm}(k') \right)}{\varepsilon_m(k') - \varepsilon_m(k)} \tag{6.125}
\]

This solution can be substituted back into Eq. (6.121). This allows the extraction of the singularity given by \( \delta(\varepsilon_n(k') - \varepsilon_m(k)) \). We have:

\[
f_{nm}(k', k) = -i \pi \delta(\varepsilon_n(k') - \varepsilon_m(k)) \times \left[ \\
\hbar \Omega_{nm} \left( V_{k' - k}^{m} \frac{f_{nm}(k) - f_{mm}(k')}{\varepsilon_m(k) - \varepsilon_m(k')} - V_{k' - k}^{n} \frac{f_{nn}(k) - f_{nn}(k')}{\varepsilon_n(k) - \varepsilon_n(k')} \right) \\
+ V_{k' - k}^{n} f_{nm}(k) - f_{nm}(k') V_{k' - k}^{m} \right] \tag{6.126}
\]
With this solution we can solve Eq.(6.119) for $f_{nm}(k)$. By assuming that $f_{nm}(k)$ is not singular, the steady-state limit reads:

$$
(\varepsilon_n(k) - \varepsilon_m(k)) f_{nm}(k) = \hbar \Omega_{nm} (f_{nm}(k) - f_{mn}(k)) - \sum_q (V_{k-q}^n f_{nm}(q, k) - f_{nm}(k, q)V_{q-k}^m).
$$

We have to compute terms $A(k) = \sum_q V_{k-q}^m f_{nm}(q, k)$ and $B(k) = \sum_q f_{nm}(k, q)V_{q-k}^m$. The singularities impose special values for the vector $q$.

Before we evaluate $A$ and $B$, we first neglect products $V^n V^m$ for $n \neq m$. We therefore assume vanishing correlations between concomitant intra-subband transitions in different subbands. In the pure phase contribution to the dephasing time (6.4.7) between a pair of subband: $|V^n - V^m| = |V^n|^2 + |V^m|^2 - 2V^n V^m$, this is to neglect the correlation term $2V^n V^m$. However the latter is significant as it reduces the linewidth. This approximation is therefore quite crude but gives simpler algebra. In the final result, the value of the broadening energy that account for the correlation term can be used for the linewidth. We make this approximation here because we do not want to account for the dynamics that such correlation terms may induce in the equations. However such terms are relevant when the dynamics of the whole heterostructure is considered. As it was shown in [5]. The actual broadening in quantum cascade lasers is found at an intermediate value between the no-correlation case (broad transitions) and the Ando's model used in (5). However the correct value requires to solve non-equilibrium Green’s functions models or density-matrix equations with the correlation term.

We use notation $\epsilon(k) \equiv \hbar^2 k^2/2m_w$ for the kinetic energy in all subband $m$ and $\hbar \Delta_{nm} = \varepsilon_n(0) - \varepsilon_m(0)$ for the detuning energies between subband-edges.

In $A(k)$ we have $\delta(\varepsilon_n(q) - \varepsilon_m(k))$ which yields the absolute value for the exchanged wave-vector of:

$$
q_- \equiv \hbar^{-1} \sqrt{2m_w(-\hbar \Delta_{nm} + \epsilon(k))},
$$

while in $B(k)$ we have $\delta(\varepsilon_n(k) - \varepsilon_m(q))$ which yields the value:

$$
q_+ \equiv \hbar^{-1} \sqrt{2m_w(\hbar \Delta_{nm} + \epsilon(k))}.
$$

For the term $A(k)$ we have the denominators: $\varepsilon_m(k) - \varepsilon_m(q_-) = \varepsilon_n(k) - \varepsilon_n(q_-) = \hbar \Delta_{nm}$. In the term $B(k)$ we have: $\varepsilon_m(k) - \varepsilon_m(q_+) = \varepsilon_n(k) - \varepsilon_n(q_+) = \hbar \Delta_{nm}$.

If we use the notation $V_{k-q}^n V_{q-k} = V_{k-q}^m V_{q-k}^m = |V_{k-q}^n|^2$, this yields:

$$
A(k) = -i\pi |V_{k-q_-}^n|^2 f_{nm}(k) + i\pi |V_{k-q_-}^n|^2 \frac{\hbar \Omega_{nm}}{\hbar \Delta_{nm}} (f_{nm}(k) - f_{nm}(q_-))
$$

$$
B(k) = i\pi |V_{k-q_+}^m|^2 f_{nm}(k) - i\pi |V_{k-q_+}^m|^2 \frac{\hbar \Omega_{nm}}{\hbar \Delta_{nm}} (f_{mn}(k) - f_{mn}(q_+))
$$

(6.130)
We introduce usual notations for the broadening energies due to the intra-subband scattering:

$$\gamma^n(k) \equiv \pi \sum_q \delta(\epsilon_n(q) - \epsilon_m(k)) |V_{k-q}^n|^2$$

$$\gamma^m(k) \equiv \pi \sum_q \delta(\epsilon_n(k) - \epsilon_m(q)) |V_{k-q}^m|^2$$  \hspace{1cm} (6.131)

We have reintroduced the sum over \( q \) wave-vectors: for \( \gamma^n \) the dirac function extracts \( q_- \), for \( \gamma^m \) it extracts \( q_+ \). If we substitute \( A(k) \) and \( B(k) \) back in Eq.(6.127), we have:

$$\hbar \Delta_{nm} f_{nm}(k) = i(\gamma^n + \gamma^m) f_{nm}(k) + \hbar \Omega_{nm} (f_{nn}(k) - f_{mm}(k)) + i\gamma^n \hbar \Omega_{nm} (\hbar \Delta_{nm})^{-1} (f_{nn}(q_-) - f_{nn}(k)) + i\gamma^m \hbar \Omega_{nm} (\hbar \Delta_{nm})^{-1} (f_{mm}(k) - f_{mm}(q_+)).$$  \hspace{1cm} (6.132)

The term \( \hbar \Omega_{nm} (f_{nn}(k) - f_{mm}(k)) \) drives the coherence with the population difference at a given wave-vector. The coupling constant is \( \hbar \Omega \), the tunneling energy. This term therefore gives the first-order approximation. The following terms drive the coherence with the populations differences evaluated between \( k \) and special \( q_\pm \) values. Their coupling constants are \( |V|/\hbar \Omega \). They are second-order terms given by the interplay between the tunneling matrix element and the potential. They are called scattering-assisted terms. They gives the second-order contribution to the current.

The current density can be calculated for two subbands by using Eq.(6.114). We assume that \( f_{nn} \) directly gives the electron concentration and therefore \( N_s = 1 \). The final result reads:

$$J = q_0 2 \hbar \Omega^2 \sum_k \frac{\gamma^1(k) \left(f_{22}(k) - f_{11}(q_+)\right) + \gamma^2(k) \left(f_{22}(q_-) - f_{11}(k)\right)}{\hbar^2 \Delta^2 + (\gamma^1(k) + \gamma^2(k))^2}. \hspace{1cm} (6.133)$$

### 6.6.3 Interpretation of the second-order current density

In the expression in Eq.(6.133), the current is no more driven by the net population difference between the two subbands. To recover the first order expression we have to replace \( q_\pm \) by \( k \).

We do this substitution. If we consider the same scattering potential in both subbands by setting \( \gamma \equiv \gamma^1 = \gamma^2 \) and if we assume \( 2\gamma/\hbar = \tau_{\perp}^{-1} \) we recover the first order approximation:

$$J = q_0 \sum_k \frac{2\Omega^2 \tau_{\perp}(k)}{1 + \Delta^2 \tau_{\perp}(k)^2} (f_{22}(k) - f_{11}(k)) \hspace{1cm} (6.134)$$

The result is similar to the one obtained in 0-dimensional system. But the origin of the dephasing time is explained: the dephasing time arises from intra-subband scattering. We
have however not considered inter-subband scattering in the model. The latter is known to decrease the dephasing time as expressed in Eq. (5.9).

The second-order current density can be interpreted by drawing the population differences $f_{22}(k) - f_{11}(q_+)$ and $f_{22}(q_-) - f_{11}(k)$ on the energy dispersion of the subbands represented with an absolute energy scale $E$. This is done in Fig. (6.18). If we write the populations as functions of the excess energy in their respective subbands, we have: $f_{22}(\epsilon(k)) - f_{11}(\hbar\Delta + \epsilon(k))$. On an absolute energy scale if simply means: $f_{22}(E) - f_{11}(E)$. A same reasoning apply for the term: $f_{22}(q_-) - f_{11}(k)$. The second-order current was examined in more details in [96]. The results obtained in this paper will be discussed after the effective transport model (7.1) have been introduced, since the latter was used to compute the main results.

The population differences are therefore evaluated at a constant energy rather than at a constant wave-vector. This results can be extended directly to subbands with different effective masses. If we insist to write the population differences in terms of wavevectors $q_{\pm}$ we need to take care that the latter map the correct energy and reproduce Fig. (6.18).

Figure 6.18: The second-order formula for the current density is illustrated. Two subbands 1 and 2 are considered with same mass. The detuning energy $\hbar\Delta$ is reported. The population-difference terms $f_{22}(k) - f_{11}(q_+)$ and $f_{22}(q_-) - f_{11}(k)$ are reported on an absolute energy scale, showing that the population difference is evaluated at a constant energy rather than at a constant wave-vector. The values of the momentum are given by: $q_{\pm}(k) = h^{-1}\sqrt{2m_w(\pm\hbar\Delta + \epsilon(k))}$, where $\epsilon(k) = \hbar^2k^2/2m_w$ is the kinetic energy.
for the case where $m_1 \neq m_2$.

A global energy scale is normal in non-equilibrium Green’s function models developed by A. Wacker where second-order current appears naturally [108]. The lesser function $G^<$ is resolved in energy. However the density matrix is connected by integration of $G^<$ on the energy. It seems that there is no proper way to define a global energy scale in the density matrix formalism and the energy conservation, for example for the current density appears through special wave-vectors. It is worth to mention that deeper calculations in the density matrix formalism, by taking into account the correlation terms for example, are very cumbersome. In this case it is rather efficient to consider more general models, like the non-equilibrium Green’s functions and perform approximations that may results in supplementary terms in the energy broadening of the current curve and of the optical gain line.

Non-parabolicity effects on the current shape

As the result can be applied directly to subbands with different masses, the implementation of the non-parabolicity effect should also be done at the level of the detuning energy. In Eq. (6.133) we have to replace the detuning energy by its $k$-dependent expression: $\hbar \Delta \rightarrow \hbar \Delta (k) = \varepsilon_2(k) - \varepsilon_1(k)$, as the two kinetic terms do not compensate. This effect becomes stronger as the electrons populate higher momentum states. We have computed the energy shift $\hbar \Delta (k) - \hbar \Delta (0)$ between the ground state and the first excited state of a single quantum well of Ga$_{0.47}$In$_{0.53}$As with a width of 4.0 nm. The detuning between the band-edges is $\hbar \Delta (0) = 165.5$ meV. We have:

$$\hbar \Delta (k) = \hbar \Delta (0) + \frac{\hbar^2}{2m_0} \left( \frac{1}{m_2^*} - \frac{1}{m_1^*} \right) k^2,$$

(6.135)

where $m_1^* \approx 0.052$ and $m_2^* \approx 0.072$. The latter expression can be parametrized by the $m_1^*$ state kinetic energy $E$, with $k = \hbar^{-1} \sqrt{2m_0 \ast m_1^* E}$. The dependence of $\hbar \Delta (E)$ is therefore linear and the slope is $\approx 0.27$. It means that for the thermal energy $k_b T$ at room temperature it gives a typical energy shift of $\approx 7$ meV. The non-parabolicity effect is therefore small on the current density curve.

We will neglect the non-parabolicity effect on the current shape when we treat effective 0-dimensional models. However when we will treat the case of second-order optical gain an energy shift of a few meV will be very relevant for matching the computed and the measured spectral gain curves, as we will discuss in (8).
6.6.4 Second-order current in a 0-dimensional effective model

Here we show how to formulate an effective 0-dimensional model for the second-order current. From the previous section we know that the population difference between subbands has to be evaluated at a constant energy and not at a constant wave-vector. If we look back to the current density between 0-dimensional states $a$ and $b$, we have:

$$j = -\frac{2\Omega^2\tau_\perp}{1 + \Delta^2\tau_\perp^2} (\rho_{bb} - \rho_{aa}).$$  \hspace{1cm} (6.136)

In order to incorporate the effect of second-order current, we can define effective coupling energies:

$$j = -\frac{2\tau_\perp \Omega}{1 + \Delta^2\tau_\perp^2} (\Omega_{b\to a}\rho_{bb} - \Omega_{a\to b}\rho_{aa})$$ \hspace{1cm} (6.137)

The effective parameters $\Omega_{a\to b}$ will depend on the population and on the subband alignment. We therefore need to introduce a parameter that accounts for electrons that effectively tunnel between subbands.

Figure 6.19: The $\sigma$ parameter is illustrated. For negative detuning energy, all electrons from subband $b$ contribute to the current. When the detuning energy becomes positive only electrons above the edge of subband $a$ contribute to the current. A similar diagram can be drawn for subband $a$. The population of subband $a$ is not shown for readability reasons.
We assume that we know the intra-subband distribution of electrons in the subbands \(a\) and \(b\) as a function of the kinetic energy in the subbands: \(f_{aa}(\epsilon)\) and \(f_{bb}(\epsilon)\). We first focus on the population from subband \(b\) that actually contributes to the current: if the subband-edge of \(b\) is above the subband-edge of \(a\), all electrons from subband \(b\) contribute to the current. However, if subband \(b\) is below subband \(a\) only the electrons above the kinetic energy given by the detuning energy \(\hbar \Delta_{ab}^0 = \varepsilon_a(0) - \varepsilon_b(0)\) will contribute to the current. This is illustrated in Fig.(6.19). We define a parameter \(\sigma_b(\Delta_{ab}^0)\) that accounts for these configurations:

\[
\sigma_b(\Delta_{ab}^0) = \theta(-\Delta_{ab}^0) + \theta(\Delta_{ab}^0) N_b^{-1} D \epsilon \int_{h\Delta_{ab}^0}^{\infty} d\epsilon f_{bb}(\epsilon). \tag{6.138}
\]

Where the unit step \(\theta(x)\) takes value 0 when \(x \leq 0\) and takes value 1 for \(x > 0\). With this definition one should avoid direct evaluation at \(x = 0\). However the function is continuous. If we consider \(x \to 0_-\) or \(x \to 0_+\) we have \(\sigma(x = 0) = 1\). The total population in subband \(b\) is given by \(N_b \equiv D \epsilon \int_0^{\infty} d\epsilon f_{bb}(\epsilon)\). The factor \(D\) being the density of states. This parameter is valid for an arbitrary (out-of-equilibrium) distribution. In a first approximation we are interested in the equilibrium Fermi-Dirac distribution \(f_{FD}(\mu, \beta, \epsilon) = (1+\exp(\beta(\epsilon - \mu)))^{-1}\), where \(\mu\) is the Fermi-level related to the total population in the subband and \(\beta = 1/k_B T\) the temperature of the distribution. We assume that \(f_{bb}(\epsilon) = f_{FD}(\epsilon)\) and we evaluate the integrals in the Eq.(6.138). We have:

\[
\sigma_b(\Delta_{ab}^0) = \theta(-\Delta_{ab}^0) + \theta(\Delta_{ab}^0) \frac{\log (1 + e^{\beta \mu_b} e^{-\beta h \Delta_{ab}^0})}{\log (1 + e^{\beta \mu_b})} \tag{6.139}
\]

where we have introduced the function \(\chi(x) = \theta(x) + \theta(-x)e^{-\beta|x|}\). We should take care of the case \(x = 0\) as explained above for the \(\sigma\) functions. We have assumed that the Fermi-level \(\mu_b\) was determined by \(N_b = \frac{D_\epsilon}{\beta} \log (1 + e^{\beta \mu_b})\). Where: \(D_\epsilon = D_0 m_b\) and \(D_0 = (\pi \hbar^2)^{-1}\), \(m_b\) being the mass of the electrons in the subband \(b\).

The \(\sigma\) function is illustrated in Fig.(6.20). As expected the fraction of available electrons decays rapidly with positive detuning energies. As electron temperatures low as 15 K cannot however be reached in actual systems (7.1), such computations are done for illustrative purpose only.

A corresponding parameter \(\sigma_a\) can be defined for subband \(a\). Finally the effective coupling energies are defined as:

\[
\overline{\Omega}_{b \to a} = \Omega \sigma_b(\hbar \Delta_{ab}^0) \quad \text{and} \quad \overline{\Omega}_{a \to b} = \Omega \sigma_a(\hbar \Delta_{ba}^0) \tag{6.140}
\]
6.6. Two subbands in the density-matrix formalism

Figure 6.20: The $\sigma$ function defined in Eq.(6.139) is plotted for a few electron temperatures. For negative detuning energies $\hbar \Delta_{ab}^0 \equiv \varepsilon_a(0) - \varepsilon_b(0)$ it is equal to unity since all electrons contribute to the current. For positive detuning energies, the fraction of electrons decays with respect to the detuning energy as only electrons with kinetic energies $\geq \hbar \Delta_{ab}^0$ contribute to the current.

With these effective parameters we can formulate a 0-dimensional model that implements the second-order current density. The latter requires in general to solve the $k$-dependent problem. The unit current density $j$ is given by Eq.(6.137).

However the 0-dimensional effective problem can only be solved in principle by self-consistent iterations. The effective coupling parameters $\sigma$ depend implicitly on the electron distributions in the subbands. Even with a thermal model with an uniform electronic temperature, we have to know $N_{a,b}$ in order to compute $\sigma_{a,b}$.

The problem is well-defined for a system with the periodic boundary conditions discussed in the previous sections. We assume initial populations $N_i^{(0)}$ for each subband, the upper index (0) being the iteration number. We are therefore able to compute the $\sigma$ parameters and the effective couplings. We get a solution $N_i^{(1)}$ that gives new effective coupling parameters. After some iterations we expect that the populations, and therefore the current, converge on a stable solution $N_i^{(\infty)}$ up to the desired accuracy.
Leakage current treated with the second-order formula and the low-density expression of the $\sigma$ parameters

In (6.6.1) we have examined the leakage current from the upper laser-state to an excited state in the injector region. The first order current predicted a leakage that reduces the population inversion by a factor $\approx 0.65$.

It is interesting to recompute the reduction of the population inversion by modeling the system with the second-order current formula. For that we have to assume that the temperature and the electron densities in subbands satisfy: $\beta \mu \ll 0$. In this regime we can perform the approximation: $\log(1 + e^{\beta \mu}) \approx e^{\beta \mu}$. The sigma parameter between two subbands $a$ and $b$, is given for subband $b$ by:

$$
\sigma_b \approx \chi(-\hbar \Delta_{ab}^0) = \theta(-\hbar \Delta_{ab}^0) + \theta(\hbar \Delta_{ab}^0) e^{-\beta|\hbar \Delta_{ab}|}.
$$

(6.141)

This approximation is very accurate for typical mid-infrared quantum cascade lasers, where the electronic densities in subbands are $\sim 10^{10}$ cm$^{-2}$. It holds ever for low temperatures as 50 K. This result does not involve the total population of the subband. The sigma parameters can be considered as a geometrical factor. This simplify calculations as solutions are obtained without self-consistent iterations.

If we consider the situation in Fig.(6.16), the detuning $\hbar \Delta_{2p}^0 \approx -25.5$ meV. The sigma parameters for both subbands $p$ and 2 can be computed. We have for $T = 70$ K:

$$
\sigma_p = 1 \text{ and } \sigma_2 \approx 0.015.
$$

(6.142)

The leak from the upper laser-state to state $p$ is therefore very weak. Regarding the extraction of carriers, the lower laser-state is in resonance with the extraction state of the injector region. Therefore: $\sigma_e \approx \sigma_u \approx 1$.

We write the rate equations previously solved for this system with the second-order current density:

$$
\begin{align*}
\dot{N}_2 &= \frac{J_0}{q_0} - \frac{N_2}{\tau_{21}} + \alpha_p N_p - \alpha_p \sigma_2 N_2 \\
\dot{N}_1 &= \frac{N_2}{\tau_{21}} + \alpha_e (N_u - N_1) \\
\dot{N}_p &= -\frac{N_p}{\tau_p} - \alpha_p N_p + \alpha_p \sigma_2 N_2 \\
\dot{N}_u &= -\frac{N_u}{\tau_u} - \alpha_e (N_u - N_1)
\end{align*}
$$

(6.143)
6.6. Two subbands in the density-matrix formalism

Where $\alpha_{p,e}$ are the coupling coefficients of the first order formula: $j = -\alpha (\rho_{bb} - \rho_{aa})$. For this system we have $\alpha_p \approx 1.02 \cdot 10^{12} \text{ s}^{-1}$ and $\alpha_e \approx 6.73 \cdot 10^{12} \text{ s}^{-1}$. And we get the solution for $R \equiv \Delta N(\alpha_p)/\Delta N(\alpha_p = 0)$:

$$R = \frac{1}{\tau_{21}} \frac{\tau_{21}(1 + \alpha_p \tau_p)}{1 + \alpha_p (\tau_p + \sigma_2 \tau_{21})} \quad (6.144)$$

We can evaluate this ratio with the lifetimes computed from the band-edges at 70 K (for LO-phonon interaction): $\tau_p \approx 0.82 \text{ ps}$ and $\tau_{21} \approx 1 \text{ ps}$. We obtain a value of: $R \approx 0.99$. It therefore means that the parasitic channel does not exist in this structure, at this electric field strength.

It illustrates how large are the errors when quantum cascade structures are modeled with first-order current densities. We therefore have implemented the effective model discussed above in transport models as discussed in the next section.

Current between ground-states in a superlattice. Solution for second-order current formula

For sake of completeness we investigate briefly here the current density between ground-states in a superlattice. Our purpose is not to investigate transport models for these systems but rather to show that the second-order current conceptually addresses the issue raised in (6.6.1). We consider the Eq.(6.133) for equally populated subbands $f_{22}(k) = f_{11}(k) = f(k)$. It yields:

$$J = q_0 \frac{2\hbar \Omega^2 \gamma}{\hbar^2 \Delta^2 + (2\gamma)^2} \sum_k \left[ f(q_-) - f(q_+) \right]. \quad (6.145)$$

The above expression models the current between adjacent periods. We implicitly assume that the system has thick enough barriers in order to be accurately described in the sequential tunneling picture. We have assumed $\gamma \equiv \gamma^1 = \gamma^2$ which holds in this system. To simplify the expression, we have removed the $k$-dependence of the broadening energy $\gamma$.

From the above algebraic expression and from the previous paragraphs, we know that a non-zero current is obtained between the ground-states. To give numbers from the above expression, we assume a thermal distribution in the ground-state. Without giving too much details (we perform direct integration of the distribution), we have:

$$\sum_k \left[ f(q_-) - f(q_+) \right] = \frac{m_w}{\beta \pi \hbar^2} \log \left( \frac{1 + e^{\beta \mu \chi(h\Delta)}}{1 + e^{\beta \mu \chi(-h\Delta)}} \right) \quad (6.146)$$
Figure 6.21: (a) The population difference $\sum_k \left[ f(q_-) - f(q_+) \right]$ is shown against the detuning energy between adjacent ground-states of a superlattice. The sheet carrier density is $1 \cdot 10^{10}$ cm$^{-2}$ and the effective mass is $m_w \approx 0.042 m_0$ for Ga$_{0.47}$In$_{0.53}$As wells. For low temperatures the population difference varies rapidly and reaches the sheet carrier density. (b) The current density is plotted against the detuning energy, for a dephasing time of 0.2 ps and a coupling energy of 3 meV. The current has a dispersive shape as expected and contrary to the first-order approximation. The peak current decreases with the temperature.
Where \( \chi(x) = \theta(x) + \theta(-x)e^{-\beta|x|} \) was defined previously. The effective mass is typically the mass in the wells \( m_w \approx 0.042 \, m_0 \) for Ga\(_{0.47}\)In\(_{0.53}\)As.

If we assume a sheet density of \( 1 \cdot 10^{10} \) cm\(^{-2}\), we obtain the results shown in Fig.(6.21) for various electronic temperatures. However temperatures low as 15 K cannot be reached in actual systems (7.1). The dephasing time \( \tau_{\perp}^{-1} = 2\gamma/\hbar \) was chosen to \( \tau_{\perp} \approx 0.2 \) ps and the coupling energy is equal to 3 meV.

The same current density has been modeled by using non-equilibrium Green's functions [110] in Eq.(79). The result is reported below:

\[
J_{m\rightarrow n}^{\mu \rightarrow \nu} = \frac{2e}{2\pi \hbar A} \sum_{k,k'} |H_{m,k,m,k'}^{\mu,\nu}|^2 \int dE \, A_{m}^{\mu}(k,E)A_{n}^{\nu}(k',E) \times \left[ n_F(E - e\phi_m - \mu_m) - n_F(E - e\phi_n - \mu_n) \right],
\]

where \( e = -q_0 \) and \( A \) is the normalization area of the in-plane Bloch waves. The indices \( m,n,\ldots \) denote wells and the indices \( \mu,\nu,\ldots \) denote subbands in a well. The function \( n_F(E) \) is the Fermi-Dirac distribution. The energy \( \mu_m \) is the Fermi-level associated to well \( m \) and \( \phi_m \) is the potential drop in well \( m \) caused by a constant and uniform electric field applied on the structure. The squared matrix element \( |H_{m,k,m,k'}^{\mu,\nu}|^2 \) give the coupling energy between state \( |\mu, k\rangle \) in well \( m \) and state \( |\nu, k'\rangle \) in well \( n \). The functions \( A_{m}^{\mu}(k,E) \) are spectral functions. They give the probability-amplitude density that an electron in well \( m \) and state \( \mu \) has an energy \( E \) with an in-plane momentum \( k \). In the Fermi’s Golden rule approximation the product \( A_{m}^{\mu}A_{n}^{\nu} \) is replaced by a dirac function imposing a fixed relation between energy and momentum.

The Eq.(6.147) and Eq.(6.145) are equivalent provided that we consider the current between ground-states \( \mu = \nu = 0 \) and adjacent wells: \( |m - n| = 1 \). The coupling energy is given by: \( |H|^2 = \hbar \Omega \). Regarding the product of spectral functions, the latter is a Lorentzian in a standard approximation. In Eq.(6.147) the natural energy scale in Green’s functions formalism is apparent. In particular the thermal intra-subband distributions are evaluated at a fixed energy \( E \).
Chapter 7

Towards an effective transport model

In this section we develop an effective transport model in the density-matrix formalism. It fulfills the basic requirements discussed in (6), by implementing sequential resonant-tunneling at coupling barriers and the second-order expression for the current by effective coupling energies. This enables the computation of a kind of \textit{ab initio} current-voltage curves as there is no adjustable parameters, the choice of the coupling barriers being excepted. This is an important limitation of the model: the quantum effects are modeled locally (at a barrier) and the transport is assumed incoherent elsewhere.

The model is numerically applied to the computation of current-voltage curves for mid-infrared quantum cascade structures [96]. An excellent agreement is found for the material system $\text{Ga}_{0.47}\text{In}_{0.53}\text{As} / \text{Al}_{0.48}\text{In}_{0.52}\text{As}$ lattice matched on InP.

We then discuss the competing models (7.4) found in the litterature. We pay a special attention to the quantum model developed by A. Wacker [64]. It gives us a conceptual validation of our effective model: as shown in [63], quantum cascade structures basically rely on coherent transport.

The various semiclassical models, solved using Monte-Carlo techniques [22, 49, 52, 53, 59], are also discussed although the latter may be considered in the class of scattering models [24]. In particular the hybrid model [14] from Q. Hu group for far-infrared structures is discussed. The latter is found very relevant although it is accurate at the injection resonance only, as it uses a first-order expression for the current density.
Chapter 7. Towards an effective transport model

7.1 Formal derivation of an effective transport model

In this section we formulate an effective numerical model that rely on resonant tunneling at fixed barriers in quantum cascade structures and incoherent scattering elsewhere. Before we perform calculations, we clarify the application of the model to actual quantum cascade structures.

7.1.1 Foreword on the basic physics of the model

Physical limitations of standard zero-dimensional scattering models We aim to develop an effective 0-dimensional transport model for simulation of quantum cascade lasers. In order to avoid current instabilities in scattering models [24] that rely on wavefunctions with a large coherence length (6.3), we assume that the coherence length is reduced at least to one period length of the structure. Practically we cut the latter at one barrier a least. The choice of this coupling barrier is in general not trivial and the position of the barrier changes from structure to structure. In structures based on resonant injection [9], the injection barrier is the best choice, at least at the field strength where the structure is aligned. In most of these structures, the extraction of carriers from the active to the injector region also proceed by resonant-tunneling (4.3) and the coherence of the wavefunctions should be further reduced to the active and injector regions. This is also the case of special mid-infrared (6.6.1) and far-infrared (6.5) designs based on a single-quantum-well active region previously considered in this work.

Limitation of the sequential resonant-tunneling model The structures based on diagonal designs as (6.3.3) cannot be modeled accurately by cutting the structure at the injection barrier. For this kind of structures scattering models give more accurate results for moderated field strengths.

This is due to the limitation of the resonant-tunneling model. We consider the second-order model the unit current $j$ is defined in Eq.(6.137) by:

$$j = -\frac{2\tau_\perp \Omega}{1 + \Delta^2 \tau_\perp^2} (\overline{\rho}_{b\rightarrow a} - \overline{\rho}_{a\rightarrow b}).$$

The resonant-tunneling model is efficient only around the resonance and for: $\Omega \sim \tau_\perp$, where $\Omega$ is the coupling frequency and $\tau_\perp$ a typical value of the dephasing time. If the quantum cascade structure is designed to work with large detuning energies, like in the diagonal structures, the scattering models are more efficient, as we are very far from the resonance condition. These structure can be however simulated by resonant-tunneling models by
choosing the barrier inside the injector region where detuning frequencies satisfy: $\Delta \sim \tau_\perp$ and where one expects that $\Omega \sim \tau_\perp$.

The automatic choice of the coupling barrier has been considered, but the present algorithms reported in this work have failed to give satisfactory results. Dynamical models that select for each pair of states the most accurate model (resonant tunneling or incoherent scattering) have been also investigated in the early development of the present model, following a discussion between A. Wacker and the author. However the complexity of the model would have increased and the author at the time have preferred to simplify the implementation by choosing a fixed barrier. The dynamic selection of a proper transport model is however again considered in automatic optimization of structures (8.5).

Finally it is worth to mention that the model developed here has been tested accurately on mid-infrared designs only. In the far-infrared the implemented scattering sources (5) are not sufficient. The typical computed dephasing time is far too long and the current-voltage curves are therefore unstable.

The ”effective” (measured on the current-voltage curve) dephasing time can be shortened by considering inhomogeneous broadening due to interface roughness (6.4.7). However these effects have not been investigated numerically and we are not able to provide more results.

We have not tested the model with a fixed pure phase contribution [14]. The latter would have smooth the resonances and probably stabilized the structure. This adjustable parameter would have however broken the ab initio way of modeling. However far-infrared structures could have probably be simulated at satisfactory basic level.

In this work we focus on mid-infrared structures, except for the toy model developed in the far-infrared [83]. The model was pushed to its limits with success in automatic optimization of mid-infrared structures (8).

### 7.1.2 Density matrix revisited: Liouville operator and scattering

The basic foundations of our effective model is the density matrix with dissipation. The dissipation acts on the populations by incoherent scattering between the localized states but it also acts on the coherences by dephasing times. In the previous section we already have formulated these type of models in the study of a leakage channel in mid-infrared structures (6.6.1-6.6.4) and in the simulation of a far-infrared structure (6.5).

In the introduction of this section we already have mentioned the extension of the model to several coupling barriers, like the injection and the extraction barriers. For clarity we will
first consider only one coupling barrier (the injection barrier) and we extend afterwards
the model to several barriers.

Previously we have shown that the dephasing time and scattering cannot appear easily
at the level of the hamiltonian. The commutator \([H, \rho]\) does not give rise to the expected
dephasing time for the coherences and the scattering rates for the populations. This is
a half-truth. Here we have to clarify the structure of the hamiltonian. We can write:
\[ H = H_C + H_{\text{scatt}}, \]
where \(H_C\) is the band-structure part that contains the potential and
the kinetic operator \((H_0)\) and the coupling energies \((H_\Omega)\) that achieve coherent coupling
between localized states. The contribution \(H_{\text{scatt}}\) contains the interaction of the electrons
with static potentials (interface roughness, alloy disorder, etc.) and various quasi-particles
(phonons, other electrons, etc.). We have shown in (6.6) that the inclusion of an elastic
intra-subband scattering potential \(V\) yields the expected dephasing time (arising from the
scattering potential) in Eq.(6.132) for the steady-state solution of the coherences. In this
model the contribution to the equation of motion was no more than \([H_C + V, \rho]\). The model
was simple enough to be solved. For more complicated interactions like phonon scattering
it is far more efficient to consider the many-body system and to work in the framework
of non-equilibrium Green’s functions where scattering is incorporated in the dynamics of
the correlation function.

For the approximations we are considering here, like scattering (for the phase and the
populations) in the Fermi Golden rule, the problem cannot be formulated at the level of
the hamiltonian. We have to consider the part \(H_C\) and \(H_{\text{scatt}}\) independently. The modeling
of the coherent coupling between eigen-states of \(H_0\) is achieved by the term \([H_C, \rho]\). The
scattering (assuming our approximation level) cannot be treated in this way. Here we
first assume coherent propagation of wavefunctions on a localized potential by isolating
a period between infinite lateral barriers. Then we compute tight-binding parameters for
the coherent coupling of the localized states between adjacent periods. The population
relaxation is computed by incoherent scattering between states inside a period. The phase
relaxation is computed between states belonging to adjacent periods.

In order to model this configuration we need to add scattering terms to the time evolution
equation of the density matrix. This requires to consider super-operators that couple
directly the density matrix elements between them (we do not consider inhomogeneous
terms in the master equation) without imposing the special commutator structure given by
\([H, \rho]\) (here \(H\) is an arbitrary hamiltonian without the previous structural assumptions).
The equation of motion of the density matrix can be rewritten as:
\[
\frac{i\hbar}{d} \frac{d\rho}{dt} = [H, \rho] \equiv \mathcal{L}_0 \cdot \rho.
\]
(7.2)
The super-operator \(\mathcal{L}_0\) is the Liouville operator or Liouvillian. The latter is formally de-
7.1. Formal derivation of an effective transport model

fined by $\mathcal{L}_0 = [H, \cdot]$. Before choosing a particular representation and implement effective models, we consider briefly some properties of this operator. We will get insight by expressing it as a matrix through an isomorphism discussed below. It is however worth to mention that the matrix representation is only useful when we consider small systems, or on the opposite side if we want to formulate large systems in a numerical model and use the usual linear algebra theorems and methods, like diagonalisation.

On the mathematical side, $\mathcal{E}_0$ is the Hilbert space of the problem (the existence of the latter is however a mathematical abyss in many cases), we can define the space of the observables (hermitian linear operators) on $\mathcal{E}_0$. The latter is written here $\text{Op}(\mathcal{E}_0)$. With these definitions we can define the Liouville (super-)operator. This operator lives formally in the space of the linear operators on linear operators on $\mathcal{E}_0$. We may name this space $\text{Op}(\text{Op}(\mathcal{E}_0))$. Formally we can write:

$$
\mathcal{L}_0 : \text{Op}(\mathcal{E}_0) \rightarrow \text{Op}(\mathcal{E}_0) \\
A \mapsto [-i\hbar, [H, A]]
$$

(7.3)

In a basis, the latter can be considered as a matrix of matrix, this is a four-index object $L_{abcd}$. We will use this representation later in the formulation of our transport model.

It may be more intuitive to consider isomorphisms for practical reasons. If we assume that $\mathcal{E}_0$ has a dimension $N$, then $\text{Op}(\mathcal{E}_0) \sim \mathbb{C}^{N^2}$. The space $\text{Op}(\text{Op}(\mathcal{E}_0))$ is therefore isomorphic to the space of linear operators on $\mathbb{C}^{N^2}$. Observables, like the density matrix, are represented as vectors. For an observable $A$ we may choose for example the representation: $A \sim (A_{11}, A_{12}, \ldots, A_{1N}, A_{21}, \ldots, A_{2N}, \ldots, A_{kj}, \ldots, A_{NN})$. A super-operator is therefore represented by a $N^2 \times N^2$ matrix. These enable the usage of standard linear algebra (diagonalisation, etc) on super-operators.

The simplest system usually treated in the density-matrix formalism is a two-state system with the hamiltonian considered in (6.4.1). The density matrix is therefore represented by a $2 \times 2$ matrix and the Liouville operator by a $4 \times 4$ matrix.

If we choose the isomorphism $\rho \sim (\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22})$, we can write the matrix of the Liouville operator as:

$$
\frac{\mathcal{L}_0}{i\hbar} = \begin{pmatrix}
0 & i\Omega & -i\Omega & 0 \\
-i\Omega & -i\Delta & 0 & -i\Omega \\
i\Omega & -i\Delta & 0 & i\Omega \\
0 & i\Omega & i\Omega & 0
\end{pmatrix}.
$$

(7.4)

This is nothing but the evaluation of the commutator $[H, \cdot]$ written in the basis $|L, R\rangle$. In order to lighten the notation, we use index 1 for index $L$, and index 2 for index $R$.

We have assumed previously that the coherences $\rho_{21}$ and $\rho_{12}$ decay on the typical time
scale given by $\tau_\perp$. The equations for $\rho_{ab}$ read:

$$
\dot{\rho}_{12} = -\hbar \Omega (\rho_{22} - \rho_{11}) - i \Delta \rho_{12} - \frac{\tau_{\perp}}{\tau_{1}} \rho_{12} \\
\dot{\rho}_{21} = \hbar \Omega (\rho_{22} - \rho_{11}) + i \Delta \rho_{21} - \frac{\tau_{\perp}}{\tau_{2}} \rho_{21}
$$

(7.5)

The damping of the coherences can be translated as a super-operator:

$$
\frac{\mathcal{P}}{i\hbar} = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & -\frac{\tau_{\perp}}{\tau_{1}} & 0 & 0 \\
0 & 0 & -\frac{\tau_{\perp}}{\tau_{2}} & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
$$

Similarly, the incoherent scattering between populations $\rho_{11}$ and $\rho_{22}$ are modeled with the equations:

$$
\dot{\rho}_{11} = \hbar \Omega (\rho_{21} - \rho_{12}) - \frac{\rho_{11}}{\tau_{1}} + \frac{\rho_{22}}{\tau_{2}} \\
\dot{\rho}_{22} = -\hbar \Omega (\rho_{21} - \rho_{12}) + \frac{\rho_{11}}{\tau_{1}} - \frac{\rho_{22}}{\tau_{2}}.
$$

(7.6)

The population relaxation is translated by a super-operator $\mathcal{R}$ that reads:

$$
\frac{\mathcal{R}}{i\hbar} = \begin{pmatrix}
-\frac{1}{\tau_{1}} & 0 & 0 & \frac{1}{\tau_{2}} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\frac{1}{\tau_{1}} & 0 & 0 & -\frac{1}{\tau_{2}}
\end{pmatrix}
$$

The equations of motion of the density matrix with dissipation read:

$$
i\hbar \dot{\rho} = (\mathcal{L}_0 + \mathcal{P} + \mathcal{R}) \cdot \rho \equiv \mathcal{L} \cdot \rho
$$

(7.7)

The scattering $H_{\text{scatt}}$ can therefore be modeled up to first-order transitions with additional linear super-operators $\mathcal{P}$ and $\mathcal{R}$. The stationary solution $\dot{\rho} = 0$ is given by the diagonalization of the total super-operator $\mathcal{L} \equiv \mathcal{L}_0 + \mathcal{P} + \mathcal{R}$ for a two-state model. This is nothing but solving the equations algebraically (6.4.5). However such operators can be written for larger systems and solved numerically by using standard diagonalization routines [77].

In the early days of the numerical implementation of our effective model with one coupling barrier, we have expressed large (sparse) matrices for these operators. We have solved the problem numerically for three adjacent periods coupled by resonant-tunneling with periodic boundary conditions. We have implemented optimization routines that were selecting only non-trivial equations of motion for the density-matrix elements (we were not interested in the coherences inside the period as we assumed that the coherence of these states was perfect across one isolated period). We have gained a lot of time in testing the model at the price of a large effort in the coding of the numerical implementation.
7.1. Formal derivation of an effective transport model

The solution to our problem can however be implemented in a much more easy way by looking formally to both the equations of motions and the structure of the operators. The problem can be recast into effective rate equations that are less power consuming and simpler to implement numerically. In the next paragraphs we give the detailed calculations.

First we come back to the general definition of the Liouville operator. We can write its components by choosing a basis. The equations of motion of the density matrix read:

\[ i\hbar \dot{\rho}_{ij} = \sum_{kl} \mathcal{L}_{ij,kl} \rho_{kl} \]  \hspace{1cm} (7.8)

The hermiticity of the energy operator \( H^\dagger = H \) can be translated into a condition on the components of \( \mathcal{L}_0 \):

\[ \mathcal{L}_{0,ij,kl} = -\mathcal{L}_{0,jl,ik} \]  \hspace{1cm} (7.9)

This condition is of central interest in the formulation of the problem as we have to ensure that the dephasing operator \( \mathcal{P} \) and the population-relaxation operator \( \mathcal{R} \) fulfill these requirements. If not, the conservation of the norm is not ensured and the expectation values of observables, like the current operator, will not be real numbers.

In a self-consistent problem with periodic boundary conditions the hermiticity conditions is necessary but not sufficient. For example we will have to ensure the consistency of the rate equations: the total out-scattering rate from a state have to be defined as the sum of the scattering rate from this state to all other states.

The condition in Eq.(7.9) is furthermore important as the problem needs to be well-defined at the super-operator level. As shown in quantum mechanics text-books [20] the scattering from a quantum state can be modeled by adding an imaginary part to its energy. This technique models the decay of the probability to some outer-system. However as the decay rate can be controlled by the numerical value of the imaginary part, the scattering between the states is completely out-of-control as the conservation of the norm is broken by a non-hermitian hamiltonian. A quantum system like quantum cascade lasers cannot be modeled efficiently by using this technique.

### 7.1.3 General structure of the scattering operators

In this part we give the mathematical structure of the scattering operators \( \mathcal{R} \) (for the populations) and \( \mathcal{P} \) (for the coherences). Before we examine the equations of motion of the density matrix for a general liouville operator \( \mathcal{L} \) that contains the commutator \([H_C, \cdot]\) but also scattering terms, we examine the contribution from the hamiltonian \( H_C = H_0 + H_\Omega \) only. In order to avoid more confusion in the notations we write: \( \mathcal{C} \equiv [H_C, \cdot] \). The
commutator gives a special structure to $C$:

$$C_{ij,kl} = H_{Cik} \delta_{jl} - H_{Cjl} \delta_{ik}$$  \hspace{1cm} (7.10)$$

For the present problem, we can give $H_C$ as:

$$H_{Cij} = \varepsilon_i \delta_{ij} + \hbar \Omega_{ij}.$$  \hspace{1cm} (7.11)$$

The energies $\varepsilon_i \equiv \varepsilon_i(0)$ are typically the subband-edges of the eigen-energy subbands in the confined potential for one period. The coupling energies $\hbar \Omega_{ij}$ are computed between states of adjacent periods by using a tight-binding model. The shift energies were neglected due to limitations of the model that were noticed and explained in (4.4.3). It is worth to remark that $H_C$ models several periods of the quantum cascade structure, as it contains the coupling energies between states of adjacent periods.

In Eq.(7.11) we can assume that the coupling matrix is real and therefore symmetric $\hbar \Omega_{ij} = \hbar \Omega_{ji}$. Although the tight-binding model developed in (4.4) yields asymmetric coupling energies in a two-state model we can define a coupling energy (geometric average) that comes out naturally by diagonalization. We therefore impose: $\hbar \Omega_{ij} \equiv \hbar \Omega_{ji} \equiv \hbar \Omega_{(ij)}$, where the notation $(ij)$ means that a unique value is defined for a pair of states $i$ and $j$.

For the components $C_{ij,kl}$ we have:

$$C_{ij,kl} = \varepsilon_i \delta_{ik} \delta_{jl} - \varepsilon_j \delta_{jl} \delta_{ik} + \hbar \Omega_{ik} \delta_{jl} - \hbar \Omega_{jl} \delta_{ik}.$$  \hspace{1cm} (7.12)$$

In order to define the phase relaxation ($P$) and the population relaxation ($R$) super-operators, we examine the equations of motion $\dot{\rho} = \mathcal{L} \cdot \rho$ for a general super-operator $\mathcal{L}$.

We start by considering the equations of motion for the populations $\rho_{ii}$ in Eq.(7.8) by setting $i = j$:

$$i\hbar \dot{\rho}_{ii} = \sum_{j,k} \mathcal{L}_{ii,kl} \rho_{kl} = \sum_k \mathcal{L}_{ii,kk} \rho_{kk} + \sum_{k,l} \mathcal{L}_{ii,kl} \rho_{kl}.$$  \hspace{1cm} (7.13)$$

In the last line, we have distinguished the sector of $\mathcal{L}$ that couples the populations between them $\mathcal{L}_{ii,kk}$ and the sector that couples populations to coherences $\mathcal{L}_{ii,kl}$. As expected the super-operator $\mathcal{C}$, for a collection of 0-dimensional states, does not couple the populations directly between them as: $\mathcal{C}_{ii,kk} = 0$ for the structure of the hamiltonian in Eq.(7.11).

We identify $\rho_{ii}$ with the population $N_i$ in state $i$: $\rho_{ii} \equiv N_i$. We have:

$$\dot{N}_i = \sum_k (i\hbar)^{-1} \mathcal{L}_{ii,kk} N_k + \sum_{k,l} (i\hbar)^{-1} \mathcal{L}_{ii,kl} \rho_{kl}.$$  \hspace{1cm} (7.14)$$
We can therefore identify: \( W_{k \rightarrow i} \equiv (\hbar)^{-1} \mathcal{L}_{ii,kk} \), where \( W_{k \rightarrow i} \) is the incoherent scattering rate from state \( k \) to state \( i \). The population relaxation super-operator \( \mathcal{R} \) has therefore the structure:

\[
\mathcal{R}_{ij,kl} \equiv \hbar W_{k \rightarrow i} \delta_{ij} \delta_{kl}.
\] (7.15)

As \( W_{k \rightarrow i} \) is a real number, the components \( \mathcal{R}_{ij,kl} \) satisfy the hermiticity condition given by Eq.(7.9). The proof is direct but we report it here for sake of completeness:

\[
-\mathcal{R}_{ji,lk}^* = -(-i)\hbar W_{l \rightarrow j} \delta_{ji} \delta_{lk} = \hbar W_{k \rightarrow i} \delta_{ij} \delta_{kl}.
\] (7.16)

In the last equality we have renamed the dummy indices \( l \) and \( j \) and used the symmetry of the \( \delta \) symbol.

As discussed previously, the hermiticity condition only does not ensure the conservation of the norm. It is standard in rate equations to define the total scattering rate for each state as:

\[
W_{i \rightarrow i} \equiv -\sum_{k \neq i} W_{i \rightarrow k}.
\] (7.17)

The scattering super-operator \( \mathcal{R} \) is therefore defined univocally by the scattering operator \( W \).

In Eq.(7.14) the sector \((\hbar)^{-1} \mathcal{L}_{ii,kl} \) for \( k \neq l \) represents the impact of the coherences \( \rho_{kl} \) on the dynamics of the populations \( N_i \). We can evaluate the contribution given by the coherent super-operator \( \mathcal{C} \). We have:

\[
(i\hbar)^{-1} \mathcal{C}_{ii,kl} = -i\Omega_{ik} \delta_{il} + i\Omega_{il} \delta_{ik}.
\] (7.18)

We then compute the contraction with the coherences \( \rho_{kl} \). It yields:

\[
(i\hbar)^{-1} \sum_{k,l \neq i} \mathcal{C}_{ii,kl} \rho_{kl} = \sum_{k \neq l} i (\Omega_{il} \delta_{ik} - \Omega_{ik} \delta_{il}) \rho_{kl} = \sum_{l \neq i} i \Omega_{il} \rho_{il} - \sum_{k \neq i} i \Omega_{ik} \rho_{ki} = \sum_{k \neq i} i \Omega_{ik} (\rho_{ik} - \rho_{ki}).
\] (7.19)

The terms in the last line have the structure of a quantum mechanical current (for \( q_0 = N_s = 1 \)). We cannot do this identification formally here as we need first to define the current by the expectation value \( \text{Tr}(\hat{Z} \rho) \).

The dynamics of the populations \( \rho_{ii} \) is therefore given by incoherent scattering and by coherent coupling (current densities) between pair of states.

We now turn ourselves to the equations of motion for the coherences \( \rho_{ij} \). We have for \( i \neq j \):

\[
i\hbar \dot{\rho}_{ij} = \sum_{kl} \mathcal{L}_{ij,kl} = \sum_k \mathcal{L}_{ij,kk} N_k + \mathcal{L}_{ij,ij} \rho_{ij} + \sum_{k \neq i} \sum_{l \neq j} \mathcal{L}_{ij,kl} \rho_{kl}.
\] (7.20)
In the above expression we have single out three terms. The first is driven by the populations. The coupling coefficient is given by $L_{ij,kk}$ for $i \neq j$. The second term couples $\rho_{ij}$ to itself with coefficient $L_{ij,ij}$ for $i \neq j$. Finally the last term couples $\rho_{ij}$ to all others coherences $\rho_{kl}$ by the components $L_{ij,kl}$ for $i \neq j$, $k \neq i$ and $l \neq j$. We examine below the contribution of $\mathcal{C}$ and we show where the dephasing times should be introduced.

It is worth to remark that the scattering super-operator $\mathcal{R}$ does not contribute directly to the evolution of the coherences. Its relevant components are identically zero as $\mathcal{R}_{ij,kl} \propto \delta_{ij}$ if we impose $i \neq j$.

We compute the various contributions of $\mathcal{C}$ to the dynamics of the coherences. For the first term in Eq.(7.20), we evaluate $\mathcal{C}_{ij,kk}$ directly:

$$\mathcal{C}_{ij,kk} = \hbar \Omega_{ik} \delta_{jk} - \hbar \Omega_{jk} \delta_{ik}$$  \hspace{1cm} (7.21)

These components extract the population difference $\rho_{jj} - \rho_{ii}$ (by using the symmetry $\Omega_{ij} = \Omega_{ji}$). As expected the coherences are driven by the population difference between the corresponding states.

For the second term in Eq.(7.20), we simply have $\mathcal{C}_{ij,ij} = \hbar \Delta_{ij}$ where the detuning energy is defined by: $\hbar \Delta_{ij} = \varepsilon_i - \varepsilon_j$.

The last term in Eq.(7.20) involves interactions of $\rho_{ij}$ with all other coherences. It is worth to remark that $\mathcal{C}$ does not contributes to this term. If we consider the elements $\mathcal{C}_{ij,kl} = \ldots \delta_{ik} \delta_{jl} - \ldots \delta_{ik} \delta_{jl} + \ldots \delta_{jl} - \ldots \delta_{ik}$ and we impose $k \neq l$, $k \neq i$ and $l \neq j$, the value of $\mathcal{C}$ is zero in any cases.

The question is where the depasing time should be introduced in the general super-operator $\mathcal{L}$. The dephasing time $\tau_{\perp,ij}$ act diagonally on $\rho_{ij}$. It is therefore implemented in the component $\mathcal{L}_{ij,ij}$. We can define the components of the super-operator $\mathcal{P}$ by:

$$\mathcal{P}_{ij,kl} \equiv -i \hbar \tau_{\perp,ij}^{-1} \delta_{ik} \delta_{jl}$$  \hspace{1cm} (7.22)

where $\tau_{\perp,ij}$ is a real number. This operator needs to satisfy the hermiticity condition in Eq.(7.9). We have:

$$- \mathcal{P}^{*}_{ji,lk} = (-i) \hbar \tau_{\perp,ij}^{-1} \delta_{jl} \delta_{ik} = i \hbar \tau_{\perp,ji}^{-1} \delta_{ik} \delta_{jl}$$  \hspace{1cm} (7.23)

We therefore have to assume that:

$$\tau_{\perp,ij} = \tau_{\perp,ji}.$$  \hspace{1cm} (7.24)

It means that the coherence $\rho_{ij}$ and its complex conjugate $\rho_{ij}^* = \rho_{ji}$ decay at the same rate. This condition is necessary in order to obtain real current densities. The latter being proportional to $(\rho_{ij} - \rho_{ji})$. 


It is interesting to check if the phase relaxation super-operator $\mathcal{P}$ contributes directly to the dynamics of the populations. The driving components are $\mathcal{P}_{ii,kl}$. We compute the contraction:

$$\sum_{kl} \mathcal{P}_{ii,kl} \rho_{kl} = -i \hbar \tau_{\perp,ii}^{-1} \sum_{kl} \delta_{ik} \delta_{il} \rho_{kl} = -i \hbar \tau_{\perp,ii}^{-1} \rho_{ii}. \quad (7.25)$$

The element $\tau_{\perp,ii}$ is \emph{a priori} not zero. However this term is meaningless. We have ensured the consistency of the rate equations by defining the diagonal elements of the scattering operator $W$ in Eq.(7.17). We therefore impose $\tau_{\perp,ii}^{-1} \equiv 0$ for all $i$.

In the definition of $\mathcal{P}$ we can rewrite: $\tau_{\perp,ij}^{-1} \rightarrow \tau_{\perp,(ij)}^{-1}$. This notation means explicitly that a unique value of the dephasing time is defined for the pair of states $i$ and $j$.

The scattering super-operators $\mathcal{P}$ and $\mathcal{R}$ are now defined. We have shown previously that the scattering rates can be computed at various levels of approximation (6.1). Self-consistent iterations may be required in order to obtain a stable value. The numerical values of the components of $\mathcal{R}$ are modified but not its mathematical structure, as for the operator $\mathcal{P}$.

### 7.1.4 Second-order currents in the effective Liouville model

We have shown in (6.6.1) that the current density needs to be computed with second-order terms in order to avoid large errors in the simulations of actual structures. As we have discussed in (6.6.4) we can formulate an effective model with 0-dimensional states that implements the effect of second-order terms for intra-subband thermal distributions. This is done by considering asymmetric effective coupling constants $\Omega_{ij}$.

In order to implement this model, we modify the equations of motion for the coherences. If we consider the first term in Eq.(7.20), we have:

$$\sum_{k} \mathcal{L}_{ij,ii} N_k = \mathcal{L}_{ij,ii} N_i + \mathcal{L}_{ij,jj} N_j \quad (7.26)$$

If $\mathcal{L} = \mathcal{C}$ the terms $\mathcal{C}_{ij,ii}$ and $\mathcal{C}_{ij,jj}$ are the coupling energies. The definition of the $\sigma_i(h\Delta_{ab})$ function is found in Eq.(6.139) for a Fermi-Dirac distribution. In order to avoid confusion, we use the notation: $\sigma_{i \rightarrow j} \equiv \sigma_i(h\Delta_{ij})$ for the factor that reduces the population of state $i$ in the $i \rightarrow j$ tunneling process. We consider an effective coherent super-operator $\overline{\mathcal{C}}$ defined by:

$$\overline{\mathcal{C}}_{ij,ii} \equiv \mathcal{C}_{ij,ii} \sigma_{i \rightarrow j} \quad \text{and} \quad \overline{\mathcal{C}}_{ij,jj} \equiv \mathcal{C}_{ij,jj} \sigma_{j \rightarrow i} \quad \text{for all } i, j \quad (7.27)$$

With these $\sigma$ functions the stationary density-matrix equations are solved by self-consistent iterations. It is worth for structures with a large sheet carrier density or for low (electronic)
temperatures. However as we have seen in (6.6.4) for typical parameters of quantum cascade structures the sigma functions are reduced to a geometrical factors that do not depend on the populations. However we keep the model as general as possible within the framework of the initial approximations. We therefore have implemented the population-dependent expression of the $\sigma$ functions.

It is worth to mention that the effective super-operator $\overline{C}$ has no more the structure of a commutator with a particular hamiltonian. It would be relevant to prove that its spectrum is positive defined. In all structure we have studied with this model we do not have encountered problems. Furthermore we will show in the next section that the effective model discussed here can be recast in rate equations.

### 7.1.5 Stationary solution: effective rate equations

The problem reads:

$$i\hbar \frac{d\rho}{dt} = (\overline{C} + R + P) \cdot \rho \equiv L \cdot \rho.$$  \hspace{1cm} (7.28)

The stationary solution is given by the eigen-vector associated with the $\lambda = 0$ eigen-value of the total Liouvillian $L$. The normalization is imposed by $\text{Tr}\rho = 1$. From a solution $\rho$, the current can be computed. However the definition of the current operator requires the precise knowledge of the spatial configuration of the subbands as in (6.6). This will be done for quantum cascade structures. The current will be defined at the coupling barriers.

As all operators $\overline{C}, R, P$ depend on the populations, we choose initial electron-distributions and we update them by successive iterations until a fixed accuracy is reached.

Here, we will not discuss the general problem with super-operators. In our case, the stationary solution of the problem is reduced to effective rate equations. The main equations for the populations $N_i \equiv \rho_{ii}$ and the coherences $\rho_{ij}$ read:

$$\frac{dN_i}{dt} = \sum_k W_{k\rightarrow i} N_k + \sum_{k \neq i} i\Omega_{(ik)} (\rho_{ik} - \rho_{ki})$$  \hspace{1cm} (7.29)

$$\frac{d\rho_{ij}}{dt} = i\Omega_{(ij)} (\sigma_{i\rightarrow j} N_i - \sigma_{j\rightarrow i} N_j) - i\Delta_{ij} \rho_{ij} - \tau_{\perp,(ij)}^{-1} \rho_{ij}$$  \hspace{1cm} (7.30)

From the last equation we get the equations for the sum $(\rho_{ij} + \rho_{ij})$ and the difference $(\rho_{ij} - \rho_{ij})$, the latter being proportional to the current. We have:

$$\frac{d}{dt}(\rho_{ij} + \rho_{ij}) = -i\Delta_{ij} (\rho_{ij} - \rho_{ji}) - \tau_{\perp,(ij)}^{-1} (\rho_{ij} + \rho_{ij})$$  \hspace{1cm} (7.31)

$$\frac{d}{dt}(\rho_{ij} - \rho_{ij}) = 2i\Omega_{(ij)} (\sigma_{i\rightarrow j} N_i - \sigma_{j\rightarrow i} N_j) - i\Delta_{ij} (\rho_{ij} + \rho_{ji}) - \tau_{\perp,(ij)}^{-1} (\rho_{ij} - \rho_{ij})$$  \hspace{1cm} (7.32)
We impose the stationary condition \( \dot{\rho} = 0 \). The Eq. (7.31) yields:

\[
\rho_{ij} + \rho_{ji} = -i \Delta_{ij} \tau_{\perp,(ij)} (\rho_{ij} - \rho_{ji}).
\] (7.33)

Substituted back into Eq. (7.32) we have:

\[
\rho_{ij} - \rho_{ji} = \frac{2i \Omega_{(ij)} \tau_{\perp,(ij)}^2}{1 + \Delta_{ij}^2 \tau_{\perp,(ij)}^2} (\sigma_{i\rightarrow j} N_i - \sigma_{j\rightarrow i} N_j).
\] (7.34)

This solution can be again substituted into Eq. (7.29) for the populations. This yields:

\[
0 = \sum_k W_{k\rightarrow i} N_k + \sum_{k \neq i} \frac{2 \Omega_{(ik)} \tau_{\perp,(ik)}^2}{1 + \Delta_{ik}^2 \tau_{\perp,(ik)}^2} (\sigma_{k\rightarrow i} N_k - \sigma_{i\rightarrow k} N_i).
\] (7.35)

In order to lighten the notation, we define the coherent scattering rate \( \mathcal{R}_{i\rightarrow j} \) by:

\[
\mathcal{R}_{k\rightarrow i} \equiv \frac{2 \Omega_{(ik)} \tau_{\perp,(ik)}^2}{1 + \Delta_{ik}^2 \tau_{\perp,(ik)}^2} \sigma_{k\rightarrow i}
\] (7.36)

And Eq. (7.35) read:

\[
0 = \sum_k W_{k\rightarrow i} N_k + \sum_{k \neq i} \mathcal{R}_{k\rightarrow i} N_k - N_i \sum_{k \neq i} \mathcal{R}_{i\rightarrow k}
\] (7.37)

The last term is simply the total coherent scattering rate from state \( i \) to all other states. As usually we define:

\[
\mathcal{R}_{i\rightarrow i} \equiv - \sum_{k \neq i} \mathcal{R}_{i\rightarrow k}
\] (7.38)

This enables the definition of the total rate operator \( \mathcal{W} \) which is the sum of the incoherent scattering rates \( \mathcal{W} \) and the coherent scattering rates \( \mathcal{R} \). The stationary solution is obtained by solving standard rate equations:

\[
\mathcal{W} \cdot \mathbf{N} = 0 \quad \text{with} \quad \mathcal{W} = \mathcal{W} + \mathcal{R}.
\] (7.39)

The solution is found by diagonalization of \( \mathcal{W} \). The eigen-vector associated with the eigenvalue \( \lambda = 0 \) is the stationary populations, assuming the normalization \( \sum_i N_i = N_s \) where \( N_s \) is total carrier concentration of the system. As we imposed consistent definitions for the incoherent and the coherent scattering rates, the norm is conserved. As \( \mathcal{W} = \mathcal{W} [\mathbf{N}, \Omega_{(ij)}, \Delta_{ij}, \tau_{\perp,(ij)}] \), the solution is found after self-consistent iterations as described previously.

The knowledge of \( \Omega_{(ij)} \) and of the stationary populations \( \mathbf{N} \) enables the calculation of the quantum mechanical current. However, we prefer to postpone the definition of the latter to the next section where we consider a periodic structure with one coupling barrier. The current density will be evaluated at this barrier.
Chapter 7. Towards an effective transport model

7.1.6 Periodic systems with one coupling barrier

We examine a periodic system. Each period is coupled to the adjacent periods by a coupling barrier potential. The basis is localized basis inside the period. It is computed by isolating a period between two infinite barrier at each side of the period. The coupling is achieved by a barrier potential over a finite region. The alignment of two adjacent periods is a free parameter. It depends on the width of the coupling barrier and on the strength of the applied electric field. However, if we keep in mind the example of quantum cascade structures, the coupling in achieved typically by the injection barrier and the alignment electric field corresponds to the alignment of the band-structure: resonance between the injector ground-state and the upper laser-level. Once the periods are aligned, we can determine the detuning energies between adjacent periods. A tight-binding model is then used to compute the coupling energies between each pair of states. Incoherent scattering inside the period is evaluated for various interactions \(5\) for an initial distribution of carriers inside the period. The dephasing times between states of adjacent periods is then evaluated. (it requires the knowledge of total scattering rates for each state). The coherent and incoherent scattering matrices are known and the system can be solved by iterations using Eq.\(7.39\).

In order to illustrate the method, we consider a typical quantum cascade structure \[9\]. Although coherent coupling needs to be implemented at the injection and at the extraction barrier as discussed in \(4.3\), we consider the injection barrier only. The localized basis inside the period is shown in Fig.\(7.1a\). The coupling between adjacent periods is illustrated in Fig.\(7.1b\).

Formally the implementation of periodicity requires three adjacent periods as in \(6.3\). If we define \(N_i^{(n)}\) as being the population of state \(i\) in period \(n\), we can write the equations of motion for this population as:

\[
\frac{d}{dt} N_i^{(n)} = \sum_{j \neq i} W_{j \rightarrow i}^{(n \rightarrow n)} N_j^{(n)} - N_i^{(n)} \sum_{j \neq i} W_{i \rightarrow j}^{(n \rightarrow n)} + \\
\sum_{j} \mathcal{B}_{j \rightarrow i}^{(n+1 \rightarrow n)} N_j^{(n+1)} - N_i^{(n)} \sum_{j} \mathcal{B}_{i \rightarrow j}^{(n \rightarrow n+1)} + \\
\sum_{j} \mathcal{B}_{j \rightarrow i}^{(n-1 \rightarrow n)} N_j^{(n-1)} - N_i^{(n)} \sum_{j} \mathcal{B}_{i \rightarrow j}^{(n \rightarrow n-1)} \tag{7.40}
\]

The incoherent scattering operator \(W\) acts only on populations inside the period and the coherent scattering operator \(\mathcal{B}\) couples periods between them. We may be tempted to compute the incoherent scattering between periods, however it is basically undefined as we assumed that the states in each period are localized. This additional, incoherent, channel is a very bad approximation of the scattering between extend states across the coupling
7.1. Formal derivation of an effective transport model

(a)

(b)

Figure 7.1: The coherent coupling between periods is illustrated for transport models with one coupling barrier. (a) The localized basis is shown. It was computed by isolating the period between infinite barriers. (b) The transport model is illustrated. Incoherent relaxation occurs inside the periods with matrix elements $W^{(n \rightarrow n)}_{i \rightarrow j}$. Adjacent periods are coherently coupled by coupling energies $\hbar \Omega_{ij}$. The detuning energies $\hbar \Delta_{ij}$ and the de-phasing times $\tau_{\perp (ij)}$ are also reported. These parameters enable the computation of the coherent scattering rates $R^{(n \pm 1 \rightarrow n)}_{i \rightarrow j}$ with Eq. (7.36).

barrier when the resonant tunneling model fails to describe the current for $\Omega \gg \tau_{\perp}^{-1}$. In this case one should rather consider tunnel-assisted scattering. For now we consider only coherent coupling between periods.
As for the incoherent case, the inter-period scattering matrices $R$ have the following symmetry:

$$R^{(m+s \rightarrow m)} = R^{(m \rightarrow m-s)}$$  \hspace{1cm} \text{for } m,n \text{ integers.} \tag{7.41}$$

We use this symmetry on terms that involve $n-1$ and $n$ periods. We obtain:

$$R_{j \rightarrow i}^{(n-1 \rightarrow n)} = R_{j \rightarrow i}^{(n \rightarrow n+1)} \quad \text{and} \quad R_{i \rightarrow j}^{(n \rightarrow n-1)} = R_{i \rightarrow j}^{(n+1 \rightarrow n)} \tag{7.42}$$

We rearrange Eq.(7.40) to obtain:

$$\frac{d}{dt} N_{i}^{(n)} = -N_{i}^{(n)} \left( \sum_{j \neq i} W_{i \rightarrow j}^{(n \rightarrow n)} + \sum_{j} R_{i \rightarrow j}^{(n \rightarrow n+1)} + \sum_{j} R_{i \rightarrow j}^{(n+1 \rightarrow n)} \right) +$$

$$\sum_{j \neq i} W_{j \rightarrow i}^{(n \rightarrow n)} N_{j}^{(n)} + \sum_{j} R_{j \rightarrow i}^{(n+1 \rightarrow n)} N_{j}^{(n+1)} + \sum_{j} R_{j \rightarrow i}^{(n \rightarrow n+1)} N_{j}^{(n-1)} \quad \tag{7.43}$$

We then assume a periodic solution: $N^{(m+s)} = N^{(m)} \equiv N$ for all $m$ and $s$ integers. We therefore drop the period index on the populations. The term $R_{i \rightarrow i}^{(n \rightarrow n+1)} N_{i}$ in the first line is compensated by the same term in the second line. As for the term $R_{i \rightarrow i}^{(n+1 \rightarrow n)} N_{i}$. This yields:

$$\frac{d}{dt} N_{i} = -N_{i} \sum_{j \neq i} \left( W_{i \rightarrow j}^{(n \rightarrow n)} + R_{i \rightarrow j}^{(n \rightarrow n+1)} + R_{i \rightarrow j}^{(n+1 \rightarrow n)} \right) +$$

$$\sum_{j \neq i} \left( W_{j \rightarrow i}^{(n \rightarrow n)} + R_{j \rightarrow i}^{(n+1 \rightarrow n)} + R_{j \rightarrow i}^{(n \rightarrow n+1)} \right) N_{j} \quad \tag{7.44}$$

We therefore define a scattering operator $\mathcal{W}$:

$$\mathcal{W} \equiv W^{(n \rightarrow n)} + R^{(n \rightarrow n+1)} + R^{(n+1 \rightarrow n)} \tag{7.45}$$

The diagonal elements $\mathcal{W}_{i \rightarrow i}$ are defined consistently:

$$\mathcal{W}_{i \rightarrow i} \equiv -\sum_{j \neq i} \mathcal{W}_{i \rightarrow j} \quad \tag{7.46}$$

Finally we obtain rate equations:

$$\frac{dN}{dt} = \mathcal{W} \cdot N \quad \tag{7.47}$$

This equation is non-linear since the scattering operator depends on the populations. The stationary solution is obtained by diagonalization of $\mathcal{W}$ and by self-consistent iterations.

If we look back to incoherent scattering models (6.3.3), Eq.(6.40) is very similar to the above equation if the following transformations are considered: $W^{(n+1 \rightarrow n)} \rightarrow R^{(n+1 \rightarrow n)}$ and $W^{(n \rightarrow n+1)} \rightarrow R^{(n \rightarrow n+1)}$. Although the structure is very similar the resulting dynamics will be completely different. We cannot obtain formally one as a limiting case of the other. The density matrix model was necessary in the formulation of the coherent scattering rates with the special factors $\sigma$ that account for the second-order current.
We now turn ourselves to the definition of the current density. It is defined at the coupling barrier. We can apply the definition of the current density (for electrons) used in the two-state problem. It gives: 

\[ J = q_0 \sum_{i,j'} i\Omega_{(ij')} (\rho_{j'i} - \rho_{ij'}). \]  

This is the sum of current densities between each pair of states. The stationary solution for the polarizations \( \rho_{j'i} - \rho_{ij'} \) is given by Eq.(7.34). It reads:

\[ i\Omega_{(ij')} (\rho_{j'i} - \rho_{ij'}) = \frac{2 \Omega_{(ij')}^2 \tau_{ij'}^2}{1 + \Delta_{ij'}^2 \tau_{ij'}^2} (\sigma_{j'i} N_{j'} - \sigma_{ij'} N_i) \]  

(7.49)

This yields:

\[ J = q_0 \sum_{ij'} (R_{j'i} N_{j'} - R_{ij'} N_i) \]  

(7.50)

We can identify the matrices: \( R_{j'i} \equiv R_{j'i}^{(n-1\rightarrow n)} = R_{j'i}^{(n\rightarrow n+1)} \) and \( R_{ij'} \equiv R_{ij'}^{(n\rightarrow n-1)} = R_{ij'}^{(n+1\rightarrow n)} \). The populations are the same by periodicity: \( N_{j'} = N_i \). We get:

\[ J = q_0 \sum_{i,j} \left( R_{j'i}^{(n\rightarrow n+1)} N_j - R_{ij'}^{(n+1\rightarrow n)} N_i \right) \]  

(7.51)

In the last line we have exchanged the two dummy indices in the first term. The current density is therefore given by the net coherent scattering rates between adjacent periods. If we had simply translated the results in Eq.(6.43), we would have obtained this result directly. However we have preferred to use the definition given within the density-matrix formalism.

With two isolated periods, a tight-binding model, incoherent scattering rates inside the period and dephasing times across the coupling barrier, we can predict the current-voltage curve for quantum cascade structures where coherent coupling is implemented at the injection barrier and incoherent scattering rates inside the period.

However, as discussed previously, we have to generalize this results for several coupling barriers. This is done in the next section.
7.1.7 Systems with several coupling barriers

Coherent coupling (sequential resonant tunneling) at several barriers is introduced to model the coherent extraction from the active region in structures based, for example, on the two-phonon design [9]. The period defined by the choice of the first barrier (the injector barrier) is broken into several sub-periods: typically the active region and the injector region as shown in Fig.(7.2a).

However the choice of the coupling barrier is arbitrary. In the formulation of the problem we consider \( M \) localized bases that are the eigen-solutions of \( M \) potentials. The basis \( \mathcal{B}_n \) is computed by solving the corresponding potential insulated by two infinitely thick barriers.

We consider the cases \( M \geq 3 \). The reductions of the model to \( M = 2 \), and to \( M = 1 \) cases are discussed at the end of this section.

The coupling is introduced between adjacent bases. We therefore have a sequence of bases \( \mathcal{B} = \{ \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_M \} \). States in \( \mathcal{B}_n \) and \( \mathcal{B}_{n\pm h} \), for \( n, h \) integers, are coupled by tight-binding for \( h = 1 \). We neglect coupling for \( h > 1 \). In order to achieve the periodicity of the system, the basis \( \mathcal{B}_M \) is coupled with basis \( \mathcal{B}_1 \).

The populations in basis \( \mathcal{B}_n \) are written \( N_k^{(n)} \), where \( k = 1, \ldots, D_n = \dim(\mathcal{B}_n) \). We model the populations in states of basis \( \mathcal{B}_n \). The latter is coupled to its nearest neighbors \( \mathcal{B}_{n-1} \) and \( \mathcal{B}_{n+1} \). We can directly use Eq.(7.40), reported below:

\[
\frac{d}{dt} N_i^{(n)} = \sum_{j \neq i} W_{j \rightarrow i}^{(n \rightarrow n)} N_j^{(n)} - N_i^{(n)} \sum_{j \neq i} W_{i \rightarrow j}^{(n \rightarrow n)} + \sum_{j} \mathcal{R}_{j \rightarrow i}^{(n+1 \rightarrow n)} N_j^{(n+1)} - N_i^{(n+1)} \sum_{j} \mathcal{R}_{i \rightarrow j}^{(n \rightarrow n+1)} + \sum_{j} \mathcal{R}_{j \rightarrow i}^{(n-1 \rightarrow n)} N_j^{(n-1)} - N_i^{(n-1)} \sum_{j} \mathcal{R}_{i \rightarrow j}^{(n \rightarrow n-1)}
\]

(7.52)

It is worth to point out that basis \( n - 1 \), \( n \) and \( n + 1 \) have in general different dimensions. The sums in the above equations have therefore to be limited to the appropriated dimension. The incoherent scattering operators are square matrices, but the coherent scattering operators, that couple periods between them, are not in general square matrices.

The above expression holds for any basis in \( \mathcal{B} \). We have to implement the boundary conditions correctly. The consistency of the rate equations will be ensured once a general scattering operator for all the structure is defined in Eq.(7.60).

The periodicity is implemented by considering the boundary bases in \( \mathcal{B} : \mathcal{B}_1 \) and \( \mathcal{B}_M \).

If \( n = 1 \) then the basis corresponding to \( n - 1 \) is the basis \( \mathcal{B}_M' \). The latter is the basis \( \mathcal{B}_M \).
7.1. Formal derivation of an effective transport model

Figure 7.2: Illustration of a periodic system with several coupling barriers. (a) Example on the two-phonon structure [9] where the active and injector regions are delimited by the injection and by the extraction barriers. (b) General case. A sequence $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_M\}$ of $M$ adjacent bases. The band-structure is illustrated in each basis. The boundary conditions are show with coupling of $\mathcal{B}_1$ to $\mathcal{B}_{M'}$ and coupling of $\mathcal{B}_M$ to $\mathcal{B}_{1'}$. The redundancy of boundary coherent scattering matrices is represented: $\mathcal{R}^{(M' \rightarrow 1)} = \mathcal{R}^{(M \rightarrow 1')}$ and $\mathcal{R}^{(1 \rightarrow M')} = \mathcal{R}^{(1' \rightarrow M)}$. 
shifted before basis $\mathcal{B}_1$. If $n = M$ the basis corresponding to $n + 1$ is the basis $\mathcal{B}_1'$. The latter is the basis $\mathcal{B}_1$ shifted in front of $\mathcal{B}_M$. The system is illustrated in Fig.(7.2b).

The periodicity can be implemented by computing four coherent scattering matrices. When $n = 1$, the last line of Eq.(7.52) should be modified. We have the transformations:

$$R(0 \rightarrow 1) \rightarrow R(M' \rightarrow 1) \quad \text{and} \quad R(1 \rightarrow 0) \rightarrow R(1 \rightarrow M').$$  \hspace{1cm} (7.53)

When $n = M$, the middle line of Eq.(7.52) should be modified by the transformations:

$$R(M + 1 \rightarrow M) \rightarrow R(1' \rightarrow M) \quad \text{and} \quad R(M \rightarrow M + 1) \rightarrow R(M \rightarrow 1').$$  \hspace{1cm} (7.54)

We therefore have four matrices. Two are redundant. By construction we have: $R(M' \rightarrow 1) = R(M \rightarrow 1')$ and $R(1 \rightarrow M') = R(1' \rightarrow M)$. This is illustrated in Fig.(7.2b). The periodicity is finally imposed by: $N^0 \equiv N^0'$ and $N^{M'} \equiv N^M$.

The Eq.(7.52) can be written in a block-matrix form. We first define a population vector $\mathcal{N}$ for the $M$ localized bases as:

$$\mathcal{N} = (N^{(0)}, N^{(1)}, \ldots, N^{(M)}) = (N_1^{(0)}, N_2^{(0)}, \ldots, N_{D_1}^{(0)}, N_1^{(1)}, \ldots, N_{D_2}^{(1)}, \ldots, N_1^{(M)}, \ldots, N_{D_M}^{(M)}).$$  \hspace{1cm} (7.55)

The intra-period scattering matrices $W^{1 \rightarrow 1}$, $W^{2 \rightarrow 2}$, $\ldots$, $W^{M \rightarrow M}$ have the following block-diagonal representation on basis $\mathcal{B}$:

$$
\begin{pmatrix}
W^{1 \rightarrow 1} & 0 & \ldots & 0 \\
0 & W^{2 \rightarrow 2} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & W^{M \rightarrow M}
\end{pmatrix}
$$  \hspace{1cm} (7.56)

The matrices that couple nearest-neighboring bases are represented by the following matrix for illustration purpose:

$$
\begin{pmatrix}
C_1 & A^{(2 \rightarrow 1)} & 0 & A^{(4 \rightarrow 1)} \\
A^{(1 \rightarrow 2)} & C_2 & A^{(3 \rightarrow 2)} & 0 \\
0 & A^{(2 \rightarrow 3)} & C_3 & A^{(4 \rightarrow 3)} \\
A^{(1 \rightarrow 4)} & 0 & A^{(3 \rightarrow 4)} & C_4
\end{pmatrix}
$$  \hspace{1cm} (7.57)

The diagonal is represented by matrices $C_n$. The extra-diagonal block matrices $A^{(n \rightarrow m)}$ achieve the coupling between periods. The limit case are treated by matrices $A^{(1 \rightarrow 4)}$ and $A^{(4 \rightarrow 1)}$ that close the system. If we look for example to the second band, we have the
7.1. Formal derivation of an effective transport model

The notation agrees with the size of the matrices. However, we have to remark that matrix \( A^{(1\to 2)} \), for example, has the matrix elements:

\[
\begin{pmatrix}
A^{(1\to 2)}_{1\to 1} & A^{(1\to 2)}_{1\to 2} & A^{(1\to 2)}_{1\to 3} & \ldots \\
A^{(1\to 2)}_{2\to 1} & A^{(1\to 2)}_{2\to 2} & A^{(1\to 2)}_{2\to 3} & \ldots \\
\vdots & \vdots & \ddots & \ddots \\
A^{(1\to 2)}_{S\to 1} & A^{(1\to 2)}_{S\to 2} & A^{(1\to 2)}_{S\to 3} & \ldots & 0 \\
\end{pmatrix}
\]

As introduced in Eq.(6.17), the notation \( A_{\mu\rightarrow \nu} \) gives the matrix element with transposed indices: \( A_{\nu\mu} \equiv A_{\mu\rightarrow \nu} \).

By using these notations we can write the scattering operator \( \mathcal{W} \) matrix as:

\[
\mathcal{W} = \begin{pmatrix}
W^{(1\to 1)} & R^{(2\to 1)} & 0 & \ldots & 0 & R^{(M'\to 1)} \\
R^{(1\to 2)} & W^{(2\to 2)} & R^{(3\to 2)} & 0 & \ldots & 0 \\
0 & R^{(2\to 3)} & W^{(3\to 3)} & R^{(4\to 3)} & \ldots & 0 \\
\vdots & 0 & \ldots & R^{(M-2\to M-1)} & W^{(M-1\to M-1)} & 0 \\
R^{(1\to M')} & 0 & \ldots & 0 & R^{(M-1\to M)} & W^{(M\to M)} \\
\end{pmatrix}
\]

We need to ensure the consistency of the rate equations by defining the diagonal elements \( \mathcal{W}_{J,J} \) as being the sum of each columns of the matrix:

\[
\mathcal{W}_{J,J} = - \sum_{K} \mathcal{W}_{J\to K} \equiv - \sum_{K} \mathcal{W}_{KJ}.
\]

The indices \( J, K \) are running from 1 to \( S = \sum_{i=1}^{M} D_i \).

The total population vector \( \mathcal{N} \) can be obtained for the stationary problem by diagonalization of \( \mathcal{W} \) as explained previously. The stationary equations read:

\[
\mathcal{W} \cdot \mathcal{N} = 0.
\]

We now discuss the adaptation of the model to \( M = 2 \) and \( M = 1 \) cases. We first consider the \( M = 1 \) case. To recover a one coupling barrier model it is necessary to consider at
least three identical bases $B_1 \equiv B_2 \equiv B_3$ in order to recover equations Eq.(7.44). The matrix of the scattering operator $\mathcal{W}$ is given by:

$$
\mathcal{W} = \begin{pmatrix}
W_{n \rightarrow n} & R_{(n+1 \rightarrow n)} & R_{(n \rightarrow n+1)} \\
R_{(n \rightarrow n+1)} & W_{(n \rightarrow n)} & R_{(n \rightarrow n+1)} \\
R_{(n+1 \rightarrow 1)} & R_{(n \rightarrow n+1)} & W_{(n \rightarrow n)}
\end{pmatrix}
$$

(7.63)

The total population vector is a repetition of the population of the fundamental period. The equations Eq.(7.44) are recovered by look at band 1, 2 or 3 of the above matrix.

The case $M = 2$ is a bit more subtle. We still have to consider three bases. The first and the second basis are in general different, but the third basis is a repetition of the first: $B_3 \equiv B_1$. This ensures the periodicity. The matrix $\mathcal{W}$ reads:

$$
\mathcal{W} = \begin{pmatrix}
W_{1 \rightarrow 1} & R_{(2 \rightarrow 1)} & R_{(1 \rightarrow 1)} \\
R_{(1 \rightarrow 2)} & W_{(2 \rightarrow 2)} & R_{(1 \rightarrow 2)} \\
R_{(1 \rightarrow 1)} & R_{(2 \rightarrow 1)} & W_{(1 \rightarrow 1)}
\end{pmatrix}
$$

(7.64)

With these special cases, the model is now able to compute the stationary populations for $M = 1, 2, 3, 4, \ldots$.

Regarding the current density, we may define it between each adjacent bases. The current between $(n, n+1)$ bases for $n$ integer is written $J_{(n, n+1)}$ and defined by the difference between coherent scattering matrices as in Eq.(7.51):

$$
J_{(n, n+1)} = q_0 \sum_{i} \sum_{j} (R_{i \rightarrow j}^{(n \rightarrow n+1)} - R_{i \rightarrow j}^{(n+1 \rightarrow n)}) N_i^{(n)}
$$

(7.65)

The model was developed to intrinsically yield the same current density at each coupling barrier: $J_{1,2} = J_{2,3} = \ldots = J_{M-1,M} = J_{M,1}$. We can therefore choose to evaluate it at an arbitrary coupling barrier. This holds for $M = 1, 2, 3, \ldots$.

### 7.2 Importance of the electronic temperature

With the model developed in (7.1.7) and the scattering rates computed in (5), we are able in principle to compute the current-voltage curve for a given quantum cascade structure, with fixed coupling barriers. However, an important parameter is still free. For each subband we assumed a thermal (Fermi-Dirac) intra-subband distribution of electrons. The averaged scattering rates are obtained with these distributions. The temperature of the electrons $T_e^{(j)}$ remains free, where $i$ is the subband index. It is worth to remark that
by assuming thermal distributions in subbands we implicitly assume that the electron-electron interaction is faster than the typical inter-subband lifetimes. In this case we may assume a thermal distribution in most of the subbands. However, this assumption is in general too rough as the typical scattering time of the electron-electron interaction is comparable to the transport time scale in most quantum cascade structures.

The simplest model is to set the same temperature for all subbands: \( T_e^{(i)} \equiv T_e \) for all \( i \). Then we can set \( T_e = T_{LO} \) where \( T_{LO} \) is the temperature of the LO-phonon bath. The latter is here identified with the lattice temperature \( T_L \). The lattice temperature is the measured (average) temperature of the active region. The approximation \( T_e = T_L \) is quite poor as there are direct experimental evidences [92] showing that the electronic temperature is much higher than the lattice temperature. A better model is a linear dependence between the lattice and the electronic temperatures:

\[
T_e = \alpha J + T_L
\]

(7.66)

where \( J \) is the current density in the structure and \( \alpha \) is the coupling constant between the lattice and the electrons. For a mid-infrared (9.0 \( \mu \)m) Ga\(_{0.45}\)As/\( \text{Al}_{0.55}\)Ga\(_{0.45}\)As quantum cascade laser with a resonant phonon-extraction [92] we have: \( \alpha \approx 28 \text{ K/kA·cm}^{-2} \).

However this model requires that input of an adjustable parameter. We may suspect \textit{a priori} that the latter depends on the structure details. However it is not the case. If we assume that the potential drop per period is \( \Delta E_0 = q_0 F L_p \), where \( F \) is the applied electric field and \( L_p \) the period length; then we may consider that the system is at equilibrium if, in average, one electron loses \( \Delta E_0 \) of energy when is crosses a period (Its kinetic energy should not vary in average). Regardless of the details of the structure, the value of \( \Delta E_0 \) is roughly proportional to the energy of the optical transition. We therefore expect that the electron temperature will basically be the same for structures with comparable emission wavelengths and comparable current densities.

To be more general we may consider one temperature per subband. However the resulting calculations show temperature overflows in some subbands. This occurs typically between a high and low populated subbands (typically the upper/lower laser-level) with an energy separation \( \Delta E \) larger than the LO-phonon energy \( \hbar \omega_{LO} \). The transitions from the upper subband put electrons to the lower subband with an excess kinetic energy of \( \Delta E - \hbar \omega_{LO} \gg \hbar \omega_{LO} \). Although intra-subband emission of LO-phonons will thermalize electrons, this injection of carriers at high \( k \) wave-vectors implies that the distribution of carriers is poorly represented by a thermal distribution. In a self-consistent, iterative, solution, the electronic temperature of the lower subband will diverge as the tail of the thermal distribution tries to represent hot electrons.
A better model would of course involve out-of-equilibrium distributions in subbands. These models require at least to discretize the reciprocal space in quanta equal to a multiple of the LO-phonon energy. This method [31] was successfully applied to the active region subbands of a structure based on a two-phonon design. Out-of-equilibrium distributions appear at the injection resonance. This model was not implemented in the early days of our numerical model. However the current development state of the numerical model may enable the implementation of k-space resolved distributions (up to the LO-phonon energy resolution) with a little effort.

At the time of writing the model retains for the computation of the electronic temperature is based on the work of P. Harrison [40]. A unique electronic temperature is modeled for all subbands. A detailed balance of the kinetic energy is performed for all relevant interactions. For a fixed bias, we expect that low electronic temperature (typically equal to or higher than the lattice temperature) will show a positive balance: the electrons are gaining kinetic energy in cascading. We assume that for high-enough electronic temperature, the kinetic balance will be found negative. In structures that are largely out-of-equilibrium (like for the active region subbands discussed above) we expect that the kinetic balance will converge to zero for an infinite temperature. Fortunately by imposing a unique electronic temperature, the active region subbands are averaged with all injector-states where the charge mainly resides. We therefore expect that for high, but finite, temperature, the kinetic energy balance will becomes negative. For a structure that does not show numerical instabilities, there exist only one temperature where the kinetic balance crosses zero. This is the solution we are seeking for. The model is studied in details in the next section.

### 7.2.1 The kinetic energy balance

In this section we develop a detailed kinetic energy balance for the fundamental period. The kinetic energy balance is computed for incoherent scattering inside the period and for resonant tunneling at the coupling barriers. The effect of absorption and stimulated emission will be taken into account in (8). We first focus on the incoherent scattering. We consider transitions that involve only pair of subbands (typically computed using the Fermi’s Golden Rule). The transition between two subbands is illustrated in Fig.(7.3a) for both elastic and inelastic scattering.

In we use the notations in Fig.(7.3a), elastic scattering between the initial \(i\) and final \(j\) subbands causes an increase of the kinetic energy equal to the detuning: \(\Delta K \equiv K' - K = h \Delta_{ij}\). If the detuning energy would have been negative, the kinetic energy would have been decreased. Finally if we write \(W_{\text{elastic}}^{i \rightarrow j}\) the elastic scattering rate from state \(i\) to state
7.2. Importance of the electronic temperature

![Figure 7.3: Illustration of kinetic energy variations between subbands. (a) Transition between two subbands, for elastic (green arrow) and inelastic (red arrow is emission and blue arrow is absorption). Depending on the value of the detuning, the variation of the kinetic energy may be positive or negative. The symbol $K$ refers to the excess kinetic energy in a subband. (b) Variation of the kinetic energy in second-order resonant tunneling. The current is split into a left to right current $j_{L \rightarrow R}$ and right to left current $j_{R \rightarrow L}$, as the two processes yield different variations of the kinetic energy (green and orange bars). When summed the variation of the kinetic energy is proportional to the current density between pair of states. (c) Intra-subband inelastic processes. The distribution is cooled down by emission while the absorption heats it up.](image)

$j$, and $N_i$ the population of state $i$, the total variation of the kinetic energy is given by:

$$\delta K_{\text{elastic}} = \sum_{i,j} N_i W_{i \rightarrow j}^{\text{elastic}} \left( \hbar \Delta_{ij} \right). \quad (7.67)$$

For inelastic scattering (like LO-phonon) we have to consider the emission and absorption processes separately. We assume that the interaction emit or absorb a single quantum $\hbar \omega_0$. If we look to Fig.(7.3a), the variation of the kinetic energy for the emission is: $\Delta K = h \Delta_{ij} - h \omega_0$, while for the absorption it is: $\Delta K = h \Delta_{ij} + h \omega_0$. If we assume scattering rates for both processes, the total variation of the kinetic energy is given by:

$$\delta K_{\text{inelastic}} = \delta K_{\text{em}} + \delta K_{\text{abs}} = \sum_{i,j} N_i W_{i \rightarrow j}^{\text{abs}} (h \Delta_{ij} + \hbar \omega_0) + \sum_{i,j} N_i W_{i \rightarrow j}^{\text{em}} (h \Delta_{ij} - \hbar \omega_0) \quad (7.68)$$
Chapter 7. Towards an effective transport model

We now consider resonant tunneling. In the first order formula, for parabolic subbands, the kinetic energy does not change as the wave-vector is conserved. In this case, the variation of the kinetic energy is therefore only due to non-parabolicity effects. This is not the case for the second-order formula where the tunneling processes at a constant energy. The mechanism is elastic as the energy is conserved during the process. We consider a pair of subbands coupled across a barrier. For the states on the left of the barrier we use primed indices \( i' \), for the states on the right we use \( j \). This is depicted in Fig.(7.3b).

For electrons tunneling from the left to the right we have the tunneling rate \( \mathcal{R}_{i' \to j} \) (we assume that the \( \sigma \)-factors are included in the definition of the tunneling rate). For a population \( N_{i'} \), the variation of kinetic energy is:

\[
\delta K_{L \to R} = \sum_{i', j} \hbar \Delta_{i' j} N_{i'} \mathcal{R}_{i' \to j}.
\]

For electrons tunneling from the right to the left, we have the tunneling rate \( \mathcal{R}_{j \to i'} \). For a population \( N_j \), the variation of kinetic energy is:

\[
\delta K_{R \to L} = \sum_{i', j} \hbar \Delta_{j i'} N_j \mathcal{R}_{j \to i'}.
\]

The total variation for one fundamental period is given by:

\[
\delta K_{\text{tunnel}} = \delta K_{L \to R} + \delta K_{R \to L} = \sum_{i', j} \hbar \Delta_{i' j} \left( \mathcal{R}_{i' \to j} N_{i'} - \mathcal{R}_{j \to i'} N_j \right) = \sum_{i', j} \hbar \Delta_{i' j} \frac{J_{(i' j)}}{q_0}. \tag{7.69}
\]

We have written the current \( J_{(i' j)} \) to underline that contributions in both directions are taken into account.

The last important contribution to the kinetic balance is the intra-subband scattering. It gives basically the efficiency of the intra-subband thermalization. The contribution of elastic scattering is trivial. However, the electrons are heated inside a subband by absorption and cooled by emission in an inelastic process as shown in Fig.(7.3c). If we write \( W_{i}^{\text{abs}} \) the intra-subband absorption scattering rate of subband \( i \) and \( W_{i}^{\text{em}} \) the intra-subband emission scattering rate, we can write (we keep one index for the scattering rate in order to avoid the confusion with \( W_{i \to i} \)):

\[
\delta K_{\text{intra}} = \delta K_{\text{em}} + \delta K_{\text{abs}} = \sum_{i} N_{i} W_{i}^{\text{abs}} (\hbar \omega_0) + \sum_{i} N_{i} W_{i}^{\text{em}} (-\hbar \omega_0) \tag{7.70}
\]

Finally we sum all contributions \( \delta K \) and we obtain the averaged variation of the kinetic energy in one period:

\[
\delta K = \delta K_{\text{elastic}} + \delta K_{\text{in} \ \text{elastic}} + \delta K_{\text{tunnel}} + \delta K_{\text{intra}}. \tag{7.71}
\]

The intra-subband processes can be included in the inelastic contribution. The expression for evaluating inter-subband and intra-subband scattering rates being the same.

It is worth to remark that we have developed here the calculations for a single coupling barrier. The extension to several barriers requires a little effort. The kinetic balance have
to be evaluated for each sub-periods separately (in fact this is only necessary in order to attribute an independent electronic temperature to each sub-periods). For incoherent scattering the above equations hold, but the contribution from resonant tunneling needs to be evaluated at each interface between sub-periods including the boundaries. For each coupling barrier, the left to right current needs to be considered as a contribution to the kinetic balance of the right basis, while the right-to-left is a contribution to the kinetic balance of the left basis.

Finally the temperature is varied between the lattice temperature (or the half in order to avoid numerical instabilities at low fields) where it is expected to be positive and a high temperature where it is negative. Standard algorithms are then applied in order to find the root of the kinetic energy balance.

In Fig.(7.4) the kinetic energy balance is plotted against the electronic temperature for a two-phonon design at alignment field. The lattice temperature is $T_L = 70 \text{ K}$. The temperature that corresponds to a zero kinetic energy balance is $T_e \approx 347 \text{ K}$. This shows that the electronic temperature is largely higher than the lattice temperature.

For now we do not show more computations. A detailed study of quantum cascade structures will be performed in (7.3). We first explain briefly the computational structure of

Figure 7.4: The kinetic energy balance is shown for a two-phonon structure at alignment field. The lattice temperature is 70 K and the root of the kinetic energy balance is at $T_e \approx 347 \text{ K}$.
the transport model with a self-consistent electronic temperature. This is done in the next section.

### 7.2.2 Computational structure of the model

In this section we briefly show how the effective transport model developed in the previous sections is practically implemented.

The localized bases are first computed with a self-consistent potential (4.5) assuming an initial distribution of carriers: this is usually a thermal distribution at the lattice temperature.

The tight-binding parameters are then evaluated with the model developed in (4.4).

The scattering rates, the dephasing times and the linewidths given by each non-radiative interactions are computed (5) for each subband and for a mesh in the reciprocal space. The thermal average is then computed with Fermi-Dirac distributions with an initial uniform electronic temperature $T_e$.

We are then able to compute the matrices $W^{(n \rightarrow n)}$ and $R^{(n+1 \rightarrow n)}$ required by the model in (7.1.7).

The scattering operator is kept constant but the electronic temperature is varied in order to obtain a zero kinetic energy balance as explained in (7.2.1).

We compute new populations and we update the scattering operator with them. This computation is iterated until convergence criterions are met on the populations and the electronic temperatures.

If we consider a super-self-consistent potential we compute the self-consistent contribution to the potential given by transport populations. This is embedded in a super-loop that we iterate until a convergence criterion on the eigen-energies is met.

However a majority of structures do not require a super-self computation. The latter can nevertheless be relevant for structures where some subbands play the role of charge reservoirs and are away from the dopants.

### 7.3 Current-voltage curves: model against experiment

If this section we essentially show the published results [96]. In this paper the second-order current expression was tested against the usual first-order formula. The simulations were then confronted to measurements.
We were interested by two typical mid-infrared structures. The first structure (sample N655) is based on a two-phonon design. The emission wavelength is \( \approx 8 \mu m \). The second structure is based on a single-quantum-well active region design (sample N258). The emission wavelength is \( \approx 7.4 \mu m \). Apart from their similar emission wavelength, the designs are very different as shown in Fig.(7.5). However both have two well-defined coupling barriers: the injection barrier and the extraction barrier that delimit the active region from the injector region.

We have inserted here a nice illustration from the published paper in Fig.(7.6). The latter underlines the importance of the second-order current. Although we have already addressed the problem of the current between ground-states in a superlattice (6.6.4), the Fig.(7.6b) shows the current density between a populated subband and an empty subband. In the first-order approximation we obtain the usual lorentzian already examined by Kazarinov and Suris. For the second-order current expression and negative detuning energies, the current is partially suppressed due to the misalignment of the subbands: only a fraction of the electrons can tunnel to the empty subband. For zero or positive detuning energies both models overlap perfectly as all electrons contribute to tunneling.

We have simulated the two-phonon design (sample N655) for a lattice temperature of 300 K with the first and the second current expressions independently. For the single-quantum-well design (sample N258) we have extended the simulation to three lattice temperatures 80 K, 180 K and 300 K. In this case we only show results obtained with the second-order current because the first-order simulation were unstable for a large range of current densities. The results are shown in Fig.(7.7).

Regarding the scattering mechanisms, we have implemented for both structures the computation of \( \Gamma_{\text{inter}} \) and \( \Gamma_{\text{intra}} \) broadening (5): LO-phonon (bulk with well-material parameters), alloy disorder, impurities (dopants) and interface roughness with typical [101] parameters of \( \Lambda = 9.0 \) nm, \( \Delta = 1.2 \) nm and \( \kappa = 1.5 \) nm. This latter parameter implement of correlation between interfaces in the growth direction. The value of 1.5 nm is used here only to decrease the importance of the spike in the active-well of the N258 design as shown in Fig.(7.5). If \( \kappa \) was set to zero, the spike would contribute to interface roughness as two independent interfaces. The latter being thinner than a monolayer this is not acceptable. The vertical correlation length (base on a gaussian model) remove completely this unphysical contribution for a length of 1.5 nm. However this length is too short to affect relevantly the other interfaces in the structure. Finally all the interfaces contribute independently to interface roughness scattering excepted the spike. For consistency of the simulations we have kept this parameter for the simulation of N655 sample, although it has no impact on this structure.
Figure 7.5: Each structure is shown at injection resonance. The layer sequence starts from the injection barrier and the thicknesses are in nm; roman, resp. bold, numbers indicate $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, resp. $\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ alloy, acting as well, resp. barrier material. (a) Layers: $4.3/1.7/0.9/5.4/1.1/5.3/1.2/4.7/2.2/4.3/1.5/3.8/1.6/3.4/1.8/3.0/2.1/2.8/2.5/2.7/3.2/2.7/3.6/2.5$. Underlined layers are $1.5 \times 10^{17}$ cm$^{-3}$ Si doped. Nominal sheet carrier density is $1.2 \times 10^{11}$ cm$^{-2}$. Period length is 68.3 nm, repeated 35 times. The optical transition occurs at $\approx 154$ meV. (b) Layers: $4.8/3.6/0.2/3.6/3.5/5.1/1.1/5.0/1.2/4.5/1.3/3.5/1.5/3.4/1.6/3.3/1.8/3.2/2.1/3.0/2.5/3.0/2.9/2.9$. Underlined layers are $3 \times 10^{17}$ cm$^{-3}$ Si doped. Nominal sheet carrier density is $3.03 \times 10^{11}$ cm$^{-2}$. Period length is 68.6 nm, repeated 35 times. The optical transition occurs at $\approx 167$ meV.
7.3. Current-voltage curves: model against experiment

Figure 7.6: (a) Effects of second-order contributions on tunneling between a pair of equally populated subbands. (A) When the detuning is negative, the subband edge of subband 1 is above the edge of subband 2. As tunneling conserves energy, the current-flow from 1 to 2 is greater than the current-flow from 2 to 1, yielding a negative net current. (B) When the subbands are aligned, the detuning is zero and both contributions cancel, yielding a zero net current. (C) The detuning is positive and therefore the edge of subband 2 is above the edge of subband 1, yielding a positive net current between subbands. (b) Empty subband 1. (A) The current is reduced as only a fraction of electrons can tunnel. (B) Models overlap perfectly.

From the results in Fig.(7.7) it is clear that the second-order current is in a better agreement with the experimental curves than the first-order current, for low currents as for the full dynamic range of the current. An important effect due to the second-order terms in the current is the absence of current offset at zero bias. This feature is an important validation of the model. The experimental curves for the N258 samples have a systematically higher voltage. This is due to a serial resistance of about 0.5 $\Omega$.

The model is able to reproduce experimental curves with a high confidence. An important fact is that there is no fitting parameters inside the model.
Figure 7.7: (a) Voltage-current characteristics; experimental data (full line), simulation with second-order resonant tunneling (dotted line), simulation with first-order model (dashed line). The current curves are shown both in log scale (left axis) for inspection of low currents and in linear scale (right axis) for inspection of the dynamic range. (b) Single-quantum-well current-voltage curves for three temperatures: 80 K, 180 K and 300 K. Measurements are displayed in full line. Simulations are displayed in broken line.

The success of this effective transport model on the current have given us the confidence to implement the modeling of the laser optical field intensity in order to compute the optical power. This is examined in (8).
7.4 Competing models

This section is devoted to a theoretical comparison between the effective model we have developed in the previous sections and the various models that can be found in the literature. Although many of them have been discussed in the text, it is useful to summarize here the discussions.

First of all we have to define our model in usual terms of the literature. Our model can be classified as a scattering model. The reason is that even if the problem is formulated in the density matrix formalism, it can be finally recast into effective rate equations with a special scattering operator Eq.(7.45-7.60) that however implements the second-order current derived from a pure density matrix model (6.6.4). The fact that the model was treated in the density-matrix formalism and that the current is defined with the quantum mechanical expression does not enable more quantum effects than the initially implemented ones.

Our model is not a pure density matrix model as on the one side we performed a lot of simplifications and on the other side these simplifications have forced us to introduce super-operators that do not have a commutator structure. Pure density matrix models were typically considered by T. Kuhn and F. Rossi [81]. In fact we have simply addressed the problem of the unphysical resonances (6.3.3) due to incoherent scattering between wavefunctions with an infinite coherence length [24,45–47].

It is however worth to say that there exists no obvious parameter in the model that enables to consider purely incoherent model as a limit of our model. We have simply chosen a more physical basis and we have performed first-order approximations within. However, if we choose to large spatial regions in our model, we may then be exposed to the same instabilities as in purely incoherent models.

The seminal models developed by R.C. Iotti and F. Rossi [48, 49] can be considered as being scattering models. The semi-classical model is basically rate equations resolved in the reciprocal space and solved using a Monte-Carlo technique [51]. It is however important to mention that R.C. Iotti [49] have also considered coherent injection processes [81]. However she has neglected the second-order terms. Her injection model is however relevant at the injection resonance. She found that the injection of carriers does not rely on coherent processes. However, as discussed previously, the current in quantum cascade structures is basically driven by coherent processes even if scattering is necessary for having a non-zero current. Incoherent scattering model are therefore subject to instabilities (6.3.3). Regarding interactions, she modeled the electron-electron scattering in the mean-field approximation. The scattering rates were probably over-estimated and led the author to
The work of F. Compagnone [21] is based on a scattering model solved by Monte-Carlo methods. In the latter they however include phonon modes quantized by the heterostructure. This is not an optimization of the model itself but rather a refinement of the electron-phonon interaction in quantum cascade structures.

The next major scattering model based on Monte-Carlo solutions was proposed by H. Callebaut in the group of Qing-Hu [15] for terahertz quantum cascade lasers. It emphasizes on the importance of the scattering of electrons on impurities (dopants). The model was better explained later [14] and was called a Density-Matrix Monte-Carlo model. There, H. Callebaut has implemented an incoherent scattering model inside the period of a far-infrared structure with a coherent coupling between the periods. This model is very close from our effective model. However the second-order terms in the current were neglected. The results obtained at the injection resonance (where one expect that the second-order terms vanish) are however convincing. In reading the paper it appears clearly that the author has met with the same difficulties as we. He also obtained a too long dephasing time in far-infrared structures. He solved the problem by introducing a pure phase contribution $T_2$ to the dephasing time (by keeping the notations in usage for magnetic resonance). This method was fruitful as the current-voltage curves were stabilized. The agreement with the experimental curves is interesting.

In [11] O. Bonno have set up a scattering model solved by Monte-Carlo. The simulator was optimized for terahertz quantum cascade structures with the modeling of electron-electron scattering with a targeted screening model. In the same line we can consider a few papers [65, 66]. However the use of wavefunctions that span three periods cause the usual instabilities that unfortunately largely limit practical applications.

A scattering model based on Monte-Carlo solving was implemented in [53] and applied to the calculation of parasitic channels in the (incoherent) injection of carriers in terahertz quantum cascade lasers. The model was applied to the analysis of the gain in terahertz quantum cascade structures [52]. The model was also applied to mid-infrared lasers and output power was successfully computed [68] and compared to experiment.

Regarding 0-dimensional scattering models, the work of P. Harrison group excepted, we can consider a recent application of resonant tunneling models on detectors [13].

In scattering models one has also to pay attention to the work of A. Gordon [36] who formulated a basis invariant density-matrix model for 0-dimensional states. This model is promising since it enables conceptually to implement our work in a basis invariant model. However, to our best knowledge, not too much effort was put on this model as
only electron-phonon and some kind of resonant tunneling are currently implemented. The control of the approximation level is also questionable.

We can mention here the density-matrix model developed by E. Dupont [25]. It enables the gain to be computed more accurately with localized basis. As the calculations are performed analytically, each term can be clearly identified. Our transport model would clearly benefit from a better computation of the optical gain for thin barriers.

The role of interface roughness was investigated by J. Khurgin [57] in quantum cascade structures. The analytical model enables the prediction of an optimal parameter-regime for the coherent coupling of the injector ground-state to the upper-laser state.

We now turn ourselves to pure quantum models. As we already mention in the text, the first quantum model formulated with out-of-equilibrium Green’s functions [41] was published in [64, 109]. The problem is there treated with a high generality on the quantum formulation side, although not all refined scattering mechanisms that can be found in Monte Carlo papers are effectively implemented in the model. Non-equilibrium Green’s functions models enable the calculation of transport properties (like current) in any coupling regime, as underlined in [113]. The choice of the basis is therefore no more a problem as it was the case for all other models, the model of A. Gordon excepted. This last point needs to be clarified: the Gordon’s model is invariant \textit{a posteriori} while the basis invariance in Wacker’s model is a basic property. This general model was applied to mid-infrared spectroscopy [75]. Applications to the terahertz [73] and development of screening models [74] have to be mentioned. The non-equilibrium Green’s function model have been also considered by T. Kubis [61, 62, 119].

It is worth to mention that fully quantum model are formally the most elegant way of modeling transport in quantum cascade structures. The computational time required by these models (with the approximations mentioned in [109]) is however large. Nevertheless, code optimization and computer clusters may solve partially this issue.
Chapter 8

Light interaction modeling in quantum cascade lasers

In this part we examine the interaction between quantum cascade heterostructures and the electromagnetic field. At the end we will be able to predict the output power of quantum cascade lasers (8.4.3). Finally we give an example of automatic optimization of quantum cascade structures for the wall-plug efficiency (8.5) in the mid-infrared for wavelengths near 4.3 µm.

Another aspect of this work is the investigation of gain mechanisms between subbands. From the results of H. Willenberg [115] and A. Wacker [109] we know that a pair of subband can provide optical gain without a net population inversion. The gain mechanism can be explained by second-order optical transitions in the density-matrix formalism and by spectral functions in the Non equilibrium Green’s functions formalism [111]. The model is closely related to the second-order calculation of the current density (6.6) and to the usual Bloch oscillator described by F. Bloch in usual bulk solids [10] and by L. Eskai and R. Tsu in superlattices [26]. We will obtain the second-order gain expression [115] and discuss experimental evidences in quantum cascade lasers based on single-quantum-well active regions in the mid-infrared [39,95]. Direct evidences based on a different experimental technique [79] are supporting the existence of second-order (Bloch) gain in steady-state. Moreover Bloch gain signatures were also observed in time-resolved experiments [85,86].

We complete afterwards the four-state model for the single-quantum-well far-infrared structure discussed in (6.5). Through a simple modeling of the laser field we will be able to compute the light-current characteristic. The latter shows special features (8.3.2) that are closely related to the underlying transport mechanisms based on resonant tunneling. These features were observed experimentally on the light curve, validating the theoretical predictions on the effect of resonant tunneling [83] on the population inversion.
This simple model open the way of the implementation of the laser field in the transport model derived in (7), enabling the computation of the optical power. The predictions of the model were confronted to experiment with a good agreement [94].

8.1 Interaction between subbands and the electromagnetic field

Before we investigate the subtle gain mechanisms between subbands in planar heterostructures (8.2), we first review the basics of the interactions between an electromagnetic wave and subbands in a heterostructure [8].

The interaction between a linearly polarized electromagnetic wave and the heterostructure states enables transitions between quantum states belonging to the same band (conduction-band, valence-bands) yielding intra-band transitions, or between states belonging to different bands yielding inter-band transitions. Any type of transition is subject to selection rules. The latter are derived by computing the transition rate between two quantum states of the heterostructure by using the Fermi’s Golden rule.

We consider only classical electromagnetic fields. The Golden Rule allows to compute the energy transferred to the heterostructure by absorption or by stimulated emission. However the calculation of the spontaneous emission rate requires the quantization of the electromagnetic field. The spontaneous emission is the result of the coupling of the electrons (matter system) with the vacuum state of the electromagnetic field. The spontaneous emission is omnidirectional. The latter is typically measured in luminescence experiments, where the matter system is prepared (pumped) with electrons in an excited state. It is worth to remark that for measuring spontaneous emitted light from the matter system no cavity-effect should be tuned at the transition energy, else stimulated emission occurs.

The classical interpretation of the absorption/stimulated emission yields a continuous rate of energy transferred between the electromagnetic field (a plane wave of frequency \( \omega \)) and the heterostructure. The latter in a quantum theory is interpreted as absorption or stimulated emission of photons of energy \( h\omega \). However the selection rules can be derived even if only the electron system is quantized.

8.1.1 The dipole approximation

In this section we briefly perform the derivation of the dipole approximation for the hamiltonian that models the interaction between the matter (the heterostructure in our case)
and the electromagnetic field. Here we are treating classical fields only. However the quantization can be done afterwards \[84\]. We consider a linearly polarized electromagnetic plane wave. Let the electrical component \( E(r,t) \) being defined by:

\[
E(r,t) = F \epsilon \cos(\omega t - q \cdot r)
\]  

(8.1)

where \( \omega \) is the frequency, \( q \) the propagation vector and \( \epsilon \) the polarization of the wave. The wave \( E \) satisfy the Maxwell equations in a dielectric medium \[50\]. Therefore \( \omega = c/nq \), where \( c \) is the speed of light and \( n \) the refractive index of the dielectric. We assume that we are in a linear medium and therefore the electric and magnetic components of the wave are orthogonal to the propagation vector: \( \epsilon \cdot q = 0 \).

The interaction between the matter hamiltonian and the electromagnetic field can only be expressed with the scalar \( U(r,t) \) and vector \( A(r,t) \) gauge potentials. The (first principle) minimal substitution for the atomic momentum \( p \) yields:

\[
H = \frac{1}{2m} \left( p - \frac{e}{c} A \right)^2 + eU
\]  

(8.2)

The electric and magnetic fields are deduced from a scalar \( U(r,t) \) and vector \( A(r,t) \) potentials. The Maxwell equations are invariant under gauge transformation \[50\]. We can therefore choose a gauge, the more practical here is the Radiation gauge (also called Coulomb gauge and Transverse gauge), where \( \nabla \cdot A = 0 \). In this gauge the scalar potential is directly the solution of the Poisson equation and therefore, if there are no sources in the Maxwell equations, we set \( U \equiv 0 \). This gauge choice is well-adapted to describe radiation.

The electric field is connected to the vector gauge potential \( A \) by:

\[
F = -\frac{1}{c} \frac{\partial A}{\partial t}.
\]  

(8.3)

We can therefore express the vector potential by:

\[
A(r,t) = -eF \frac{c}{2i\omega} \left( e^{i(\omega t - q \cdot r)} - e^{-i(\omega t - q \cdot r)} \right)
\]  

(8.4)

In the hamiltonian Eq.(8.2) we neglect \( A^2 \) terms by considering linear absorption only. The spin-orbit coupling with the atoms of the matter hamiltonian is also neglected. For III-V semiconductor alloys it means that we consider irradiation with an intensity and frequency so that the spin-orbit terms are small If we write the matter (the heterostructure) hamiltonian by \( H_0 \), the total hamiltonian is given by:

\[
H = H_0 + \frac{q_0}{2m_0c} (p \cdot A + A \cdot p)
\]  

(8.5)
The interacting part of the hamiltonian is time dependent. We assume that the $H_0$ part satisfy: $H_0 |\nu\rangle = \varepsilon_\nu |\nu\rangle$. We do not give all the details regarding these eigenstates here (multi-band model for the envelop function approximation). The electromagnetic field is expected to induce transitions between the eigen-states of $H_0$. In a first-order approximation, let $|i\rangle$ and $|f\rangle$ be the initial and final states. For absorption we assume $\varepsilon_f > \varepsilon_i$. The transition rate between the initial and final state is given by the Fermi’s Golden rule:

$$W_{i\rightarrow f} = \frac{2\pi}{\hbar} |\langle f | V | i \rangle|^2 \delta (\varepsilon_f - \varepsilon_i - \hbar \omega)$$  \hspace{1cm} (8.6)

We apply the rotating wave approximation [69] by keeping only the slow varying terms $\omega_{fi} - \omega$ (resonant terms for $\omega > 0$) in the time dependent perturbation [20], where $\hbar \omega_{fi} = \varepsilon_f - \varepsilon_i$. The interaction potential $V$ therefore reads:

$$V = \frac{i q_0 F}{4 m_0 \omega} (\epsilon \cdot p e^{-i q \cdot r} + e^{-i q \cdot r} \epsilon \cdot p)$$  \hspace{1cm} (8.7)

We examine the term $q \cdot r$. The latter gives the spatial variation of the interaction potential. If the wavelength ($7/n \mu m$ for the mid-infrared, $n$ being the refractive index) is greater than the typical length on which the potential of the heterostructure varies (about 1 nm), we retain only the constant term in the expansion: $e^{-i q \cdot r} = 1 - i q \cdot r + O(|q|^2)$. The interaction potential is therefore given by:

$$V = \frac{i q_0 F}{4 m_0 \omega} \epsilon \cdot p$$  \hspace{1cm} (8.8)

This potential is equivalent [84] (up to a ratio $\omega/\omega_{fi}$) to the dipole approximation given in most text-books [20]. For the latter the interaction potential is given by $H_{\text{dipole}} = -q_0 r \cdot E_0(t)$, where $E_0(t)$ is directly the electric field where the dipole approximation has been performed.

The hamiltonian $r \cdot E$ and $p \cdot A$ have the same degree of accuracy (same approximations) but the transition rates computed with the one are not exactly the same with the other. In a simple system the ratio $\omega/\omega_{fi}$ appears [84]. However, the $p \cdot A$ form is more practical to determine the selection rules on the transitions between heterostructure states.

### 8.1.2 Absorption coefficient of a heterostructure

We derive here the absorption coefficient $\alpha(\omega)$ for an arbitrary (satisfying the previous assumptions) heterostructure in the envelop function formalism. This calculation will lead to the selection rules. We do not reproduce the details of the calculation. They can be found in [8].
8.1. Interaction between subbands and the electromagnetic field

We consider the transition rate \( W_{i \rightarrow f} \) between \(|i\rangle\) and \(|f\rangle\) given in Eq.(8.6). As \( \hbar \omega_{fi} > 0 \) this represents the absorption rate. Respectively if we consider the transition from state \(|f\rangle\) to state \(|i\rangle\), the transition rate represents stimulated emission.

It is worth to remark that the squared matrix element \( |\langle f | V | i \rangle|^2 \) is the same in both cases. The coupling constant between the electrons’ system and the radiation is the same for absorption and for stimulated emission.

In order to determine the absorption coefficient, we need to compute the net power transferred between the heterostructure and electromagnetic field. We assume that states are populated with a Fermi-Dirac distribution \( F(\varepsilon) \). The effective transitions rates for absorption and stimulated emission therefore read:

\[
W_{i \rightarrow f} = W_0 F(\varepsilon_i) (1 - F(\varepsilon_f)) \quad \text{and} \quad W_{f \rightarrow i} = W_0 F(\varepsilon_f) (1 - F(\varepsilon_i)),
\]

(8.9)

where we have written the transition rate computed with the Fermi’s Golden rule: \( W_0 = \frac{2\pi}{\hbar} |\langle f | V | i \rangle|^2 \delta(\varepsilon_f - \varepsilon_i - \hbar\omega) \). The dirac functions for the absorption and stimulated emission are the same. For absorption we have the condition: \( \varepsilon_i + \hbar\omega = \varepsilon_f \), and for the stimulated emission, we have the condition: \( \varepsilon_f - \hbar\omega = \varepsilon_i \).

The energy transferred to the heterostructure \( \hbar\omega(W_{i \rightarrow f} - W_{f \rightarrow i}) \), per unit of time, is given by:

\[
P(\omega) = \frac{2\pi}{\hbar} \frac{q_0^2 F^2}{4m^2_0 \omega^2} \hbar\omega \sum_{i,f} \delta(\varepsilon_f - \varepsilon_i - \hbar\omega) |\langle f | \mathbf{e} \cdot \mathbf{p} | i \rangle|^2 \left[ F(\varepsilon_f) - F(\varepsilon_i) \right]
\]

(8.10)

This energy balance can then be connected to the absorption coefficient \( \alpha(\omega) \) of the heterostructure. The connection is made between \( P(\omega) \) and the variation of the total energy density \( \mathcal{E} \) of the electromagnetic field. We should have:

\[
P(\omega) = -\frac{d}{dt} \int_{\Omega} d^3r \mathcal{E}(\mathbf{r}, t)
\]

(8.11)

The definition of the absorption coefficient is not trivial as the translational symmetry in the direction of the heterostructure is broken. The absorption coefficient is conceptually defined by the attenuation coefficient of the intensity of plane waves with complex propagation vectors [50]. G. Bastard discusses the definition of the latter for a heterostructure. Finally, if we assume that we cannot resolve spatially where the absorption occurs, the integration volume can be defined as \( \Omega = S \mathcal{L} \) where \( S \) is the surface of the sample and \( \mathcal{L} \) a length largely greater than the typical well-width in the heterostructure. For quantum cascade lasers we may consider \( \mathcal{L} = N_p L_p \) where \( N_p \) is the number of periods and \( L_p \) the period length. This length should not be too large otherwise the plane wave approximation fails.
We express the energy density by using the Poynting vector $P$. The amplitude $F$ appearing in $P(\omega)$ can then be expressed in terms of $\mathcal{E}$. If we then assume that by using a time average $\langle \ldots \rangle$ over the period $2\pi/\omega$, we obtain the usual exponential attenuation law, we can write for a wave propagating in the direction of the heterostructure:

$$\langle \mathcal{E}(L) \rangle = \langle \mathcal{E}(0) \rangle \exp(-\alpha(\omega)L)$$ (8.12)

The interesting case is the propagation of the wave in the plane of the heterostructure (this result was intensively used for extracting the spectral gain from measurements in (8.2)). If we consider arbitrarily the $x$-direction, we have:

$$\langle \mathcal{E}(x) \rangle = \langle \mathcal{E}(0) \rangle \exp(-\alpha(\omega)x)$$ (8.13)

In this model we cannot resolve spatially the absorption coefficient as a function of $z$. However the absorption coefficient can be extracted from transmission measurements using the above equation. We have for $\alpha(\omega)$:

$$\alpha(\omega) = A \sum_{if} \frac{1}{m_0} |\mathbf{p}_{if}|^2 \delta(\varepsilon_f - \varepsilon_i - \hbar \omega) \left[ F(\varepsilon_f) - F(\varepsilon_i) \right],$$ (8.14)

where $A = 4\pi q_0^2/(ncm_0\omega\Omega)$. The average refractive index of the heterostructure being $n$. The bracket:

$$\mathbf{p}_{if} \equiv \langle i | \mathbf{p} | f \rangle$$ (8.15)

is proportional to the dipole matrix elements $q_0 \mathbf{r}_{if}$. If we had used the $\mathbf{r} \cdot \mathbf{E}$ hamiltonian we would have got the matrix elements for the position operator $\mathbf{r}_{if}$.

In the above expression we have neglected the spontaneous emission as the electromagnetic field is not quantized. This transition rate is however very weak for quantum cascade lasers. We have postponed its evaluation in (8.3).

The absorption coefficient obtained here is a first-order model. We will show in the next section that it implies the conservation of the in-plane wave-vector.

### 8.1.3 Selection rules for optical transitions in heterostructures

In the previous section, we have derived the (first-order) absorption coefficient between energy-eigen states of a heterostructure. However we do not have given any details for the latter. We now consider the envelop function approximation in the vicinity of a high symmetry point (typically $\Gamma$-point).
We consider for the initial (index \(i\)) and final (index \(f\)) subbands the wavefunctions:

\[
F_i(r) = u_{\nu_i}(r) f_i(r) \quad \text{where} \quad f_i(r) = \frac{1}{\sqrt{S}} e^{i k_{\perp} \cdot r} \chi_i(z),
\]

(8.16)

where \(i\) is the subband index and \(\nu_i\) the band index. The function \(u_{\nu_i}\) is the (atomic) rapidly varying function (periodic part of the Bloch wave) and \(f_i\) is the envelop function. As in (3) we assume the continuity with respect to \(z\) for in-plane states with quantum number \(k_{\perp}\). The function \(\chi_i(z)\) is a slow varying part that represents the wavefunction of the subband in the \(z\)-direction.

We do not report here, the relevancy of these wavefunctions in the computation of the absorption \([8]\) coefficient and we consider the selection rules only. We have:

\[
\epsilon \cdot p_{if} \equiv \epsilon \cdot \int_{\Omega} d^3r \, F_i^*(r) p F_f(r) \\
\approx \epsilon \cdot \langle u_{\nu_i} | p | u_{\nu_f} \rangle \cdot \int_{\Omega} d^3r \, f_i^*(r) f_f(r) \\
+ \delta_{\nu_i,\nu_f} \epsilon \cdot \int_{\Omega} d^3r \, f_i^*(r) p f_f(r)
\]

(8.17)

where \(\langle u_{\nu_i} | p | u_{\nu_f} \rangle = \int_{\Omega_0} u_{\nu_i}^*(r) p u_{\nu_f}(r)\). Here we have taken into account the variation of the \(u_{\nu_i}\) functions at the atomic scale and their periodicity on \(\Omega_0\) (elementary cell volume). On the contrary the envelop functions vary at the heterostructure scale. We have therefore assumed that the \(p\) operator acts independently on the atomic functions and on the envelop functions. This approximation enables a clear identification of the selection rules.

In the first term, the subbands \(f\) and \(i\) belong to different bands \(\nu_i\) and \(\nu_f\). The braket \(\langle u_{\nu_i} | p | u_{\nu_f} \rangle\) gives the selection rules for interband transitions. In the latter the overlap of the envelop functions has however to be taken into account. We will not consider these transitions here.

The second term gives intraband transitions imposed by the dirac functions \(\delta_{\nu_i,\nu_f}\). This gives the selection rules for inter-subband optical transitions with the matrix-elements \(\langle i | \epsilon \cdot p | f \rangle\). We insert the the envelop functions and we get:

\[
\langle i | \epsilon \cdot p | f \rangle = \frac{1}{S} \int_{\Omega} d^3r \, \chi_i(z) e^{-i k_{\perp} \cdot r} (\epsilon_x p_x + \epsilon_y p_y + \epsilon_z p_z) \chi_f(z) e^{i k_{\perp} \cdot r} \\
= (\epsilon_x p_x + \epsilon_y p_y) \delta_{i,f} \delta_{k_{\perp},k_{\perp}} + \epsilon_z \delta_{k'_{\perp},k_{\perp}} \int dz \chi_i^*(z) p_z \chi_f(z)
\]

(8.18)

For a polarization in the plane of the heterostructure (\(\epsilon_x\) or \(\epsilon_y\)), the in-plane wave-vector is conserved with \(\delta_{k'_{\perp},k_{\perp}}\) and there is no transitions between subbands: \(\delta_{i,f}\). This implies: \(\epsilon_i = \epsilon_f = \hbar \omega = 0\). As there is no absorption, these transitions are forbidden.

The only non-trivial term is for a polarization in the \(z\)-direction (non-zero \(\epsilon_z\)), where a transition can occur between the subbands and therefore electromagnetic energy can be
absorbed or emitted (through stimulated emission). The in-plane wave-vector is conserved with $\delta k_{\perp, \perp}$. This is the case if we keep only first-order terms.

The electric field needs therefore to be parallel to the growth direction in order to cause transitions between subbands. As the medium is linear, the electromagnetic wave has a propagation vector parallel to the plane of the heterostructure. This particular field-configuration has an important impact on the design of laser cavities. The geometry of the latter need to confine an electromagnetic field with this special polarization.

The absorption of electromagnetic energy with an electric field in the plane of the heterostructure has an important impact on the calculation of the losses in the optical cavity. However as shown by selection rules, the free carrier absorption cannot exist in the heterostructure, except if we consider a scattering potential (impurities, phonons) that enables the conservation of energy. There is no optical intra-subband absorption which would have been the analogue of free-carrier absorption in bulk materials.

This was recently examined in details in [112]. The limits between free carrier absorption in bulk materials, absorption in superlattices and inter-subband absorption in heterostructures are performed on the basis of an analytical argument and numerical simulations achieved with a model based on non-equilibrium Green’s functions.

In (8.2) we will derive second-order gain mechanisms in the density-matrix formalism. It is worth to say that the analogue of the free carrier absorption in bulk material is at the same time obtained. By assuming a static intra-subband scattering potential $V$ in the plane of the heterostructure, the energy can be conserved by intra-subband scattering and scattering-assisted optical absorption can occur between subbands. When the second-order gain formula between two subbands, is generalized to all subbands in the heterostructure, the losses due to free-carrier absorption can be computed in the heterostructure.

The free carrier absorption is included in inter-subband transitions where intra-subband scattering occurs. The wave-vector is therefore not conserved due to the scattering potential. However absorption at optical frequency can occur as the energy conservation is restored.

When inter-subband transitions are allowed (the electric field is parallel to the growth direction), the integral $\int dz \chi_i^* p_z \chi_h^*$ gives the strength of the coupling. This is in fact the matrix elements of the dipole operator. However in place of this formula, we will use the oscillator strength and dipoles computed in the effective two band model [89] in Eq.(3.36).
8.2 Gain mechanisms in heterostructures

In this section we develop gain models between pair of subbands. We first review the first-order gain where the in-plane momentum is conserved. Then we examine different kinds of broadening sources. We first consider homogenous broadening where the spectral distribution is lorentzian. We consider briefly the inhomogenous broadening given by long range disorder (inhomogeneity of the heterostructure for length scales comparable to the period length). The latter is simply modeled by a convolution of the homogeneous gain shape with a gaussian distribution.

The transition between homogeneous and inhomogeneous broadening is however introduced ad hoc. A deeper physical model would also include higher-order (perturbation) terms in the modeling of interactions between electrons and scattering potentials. These terms lead to localization of electrons in the plane of the heterostructure (with interface roughness typically). This source of inhomogeneous broadening seems to need to be distinguished from long-range disorder in the growth direction. Both occur at the same time. If a random distribution is considered in both cases, both can be described my a same model. Here we treat long-range disorder by using a convolution of the homogeneous spectral distribution with a gaussian. We assume that interface roughness leads both to an homogeneous contribution by inter-subband scattering and to an inhomogeneous contribution that can be modeled by random long-range disorder. However we are not considering non-random, systematic (drift), growth imperfections that also leads to inhomogeneous broadening.

We finally consider the effects of non-parabolicity on first-order transitions. We develop afterwards a gain model that keeps second-order terms [115] (tunneling and intra-subband scattering) and show that it predicts gain without a net population inversion between subbands. The spectral gain is found dispersive-shaped around the central frequency. We then show that the dispersive shaped gain is very sensible to non-parabolicity. We need to model the gain shape accurately if we want to compare theoretical predictions to spectrally resolved measurements of the net modal gain in single-quantum-well mid-infrared structures [95].

The experimental evidences for Bloch gain in mid-infrared quantum cascade lasers are then discussed.
8.2.1 First-order gain model

In this section we discuss the first-order gain model where the in-plane wave-vector is conserved. The model itself is quite usual and we therefore rapidly extend the discussion to the origin of broadening. However, it would have been possible to first derive the second-order gain expression and then to show how the usual (first-order) approximation is recovered. This latter calculation will however be done in the appropriated section (8.2.4).

As shown by the expression for inter-subband absorption in Eq.(8.14) the gain (negative absorption) is given by the net population difference between subbands. The central line is there given by a dirac function \( \delta(\varepsilon_f - \varepsilon_i - \hbar\omega) \) as not broadening mechanism was considered.

Without a lot of details, we can deduce the homogeneous (with a lifetime) spectral gain shape of the first order model with simple density-matrix equations [69]. If we consider a 0-dimensional problem between an upper state \( a \) at energy \( \varepsilon_a \) and a lower state \( b \) at energy \( \varepsilon_b \), assuming \( \varepsilon_a - \varepsilon_b > 0 \), the equation of motion for the populations \( C_a(t) \) and \( C_b(t) \) are given by:

\[
\dot{C}_a = -\frac{1}{2} \left( \hbar \frac{1}{\gamma_a} + i\delta \right) + \frac{i}{2} R_0 C_b \\
\dot{C}_b = -\frac{1}{2} \left( \hbar \frac{1}{\gamma_b} - i\delta \right) + \frac{i}{2} R_0 C_a
\]  

(8.19)

The matrix element \( R_0 \equiv |\langle a | V | b \rangle| \) is the Rabbi flopping frequency. The operator \( V \) is the dipole hamiltonian (8.1.1), which reads: \( R_0 = \mathcal{P} E_0 / \hbar \), where \( E_0 \) is the magnitude of the electric field and \( \mathcal{P} \) is the dipole matrix (we assume that the latter does not vanish between states \( a \) and \( b \)).

The detuning energy is defined by: \( \hbar \delta = \hbar\omega - h\nu \), where \( \nu \) is the frequency of the electromagnetic plane wave and \( \hbar\omega = \varepsilon_a - \varepsilon_b \).

The energies \( \gamma_a,b \) have the effect of damping coefficients. They represent the lifetimes of the states. We assume that the electrons decay from states \( a \) and \( b \) to some larger system. The latter is not described here.

If we solve this system to the lowest order in the time perturbation series [20, 69], we get for the stimulated absorption\(^1\) (transition probability to state \( a \)):

\[
|C_a(t)|^2 \approx \frac{1}{4} R_0^2 \hbar^{-1} \gamma \left[ \frac{\sin(\frac{1}{2}(\omega - \nu)t)}{\frac{1}{2}(\omega - \nu)} \right]^2.
\]  

(8.20)

We have set for simplicity \( \gamma_a = \gamma_b \equiv \gamma \). It shows that the electrons oscillate between state

\(^1\)This is not the usual name of this term, but it states explicitly that absorption occurs in the presence of an electromagnetic field. There is of course no spontaneous absorption.
8.2. Gain mechanisms in heterostructures

$a$ and $b$ and decay with an exponential envelop to some larger system. The normalization of the two states is of course not ensured.

This calculation enables to give an interesting interpretation for this two-state system. The spectral distribution of the stimulated emission can be computed by preparing an electron in state $a$ and then computing the total probability that it decays from state $b$, by spontaneous emission, to the rest of the system. The rates $\gamma_{a,b}/\hbar$ are therefore interpreted as spontaneous emission rates to the rest of system (but not between states $a$ and $b$). This gives the spectral distribution of stimulated emission since we have forced implicitly the system to first decay by stimulated emission from state $a$ to state $b$ and then to decay to the rest of the system by spontaneous emission. We then integrate the total probability $P_s$ of decay from level $b$ by spontaneous emission. The element $|C_b(\tau)|^2$ being the probability per unit of time that the electron decays from state $b$, we have:

$$P_s = \frac{\gamma_b}{\hbar} \int_0^\infty d\tau |C_b(\tau)|^2. \quad (8.21)$$

If we let $|C_a(0)|^2 = 1$, then $|C_b(\tau)|^2$ is given by Eq.(8.20). With $\gamma_a = \gamma_b \equiv \gamma$, we get:

$$P_s(\nu) = \frac{1}{2} \frac{R_0^2}{(\omega - \nu)^2 + (\gamma/\hbar)^2}. \quad (8.22)$$

The spectral distribution of the gain between 0-dimensional states if therefore given by a Lorentzian of frequency-width $\gamma/\hbar$. We can therefore conclude that the homogeneous broadening, related here to lifetime, has a Lorentzian distribution. This is a lorentzian resonance centered at the resonant energy $\varepsilon_a - \varepsilon_b$. The dirac function has therefore to be replaced by a lorentzian in Eq.(8.14) for states with a finite lifetime, which is of course always the case.

This model can be extended to subbands directly. We first assume that the subbands have the same mass and that for each in-plane wave-vector $k$ the lifetime-broadening $\gamma$ is constant. As the first-order model conserves the wave-vector in an optical transition, for every $k$ state we have identical two-state (0-dimensional) systems. If we consider only the initial $i$ and final $f$ subbands in the heterostructure, we obtain the absorption cross-section:

$$\alpha(\omega) = \frac{q_0^2 |z_{if}|^2 E_{fi}^2}{L_p \varepsilon_0 n_r c \hbar^2 \omega (E_{fi} - \hbar \omega)^2 + \gamma^2}. \quad (8.23)$$

Where $z_{if}$ is the dipole element between the subbands (computed in the effective two-band model) and $E_{fi} = \varepsilon_f(0) - \varepsilon_i(0)$ is the energy detuning between the subband-edges. The refractive index is given by $n_r$ (it is in fact the refractive index of the optical mode in the

\[ ^2 \text{We have expressed here } \alpha(\omega) \text{ as the limit of the second-order gain model (8.2.4).} \]
Chapter 8. Light interaction modeling in quantum cascade lasers

cavity). The period length $L_p$ has been introduced to convert the volume concentrations of carriers $N_{i,f}$ into sheet carrier densities $n_{i,f}$. The homogeneous broadening $\gamma$ consists in the lifetime broadening and in the phase broadening (we haven’t discuss the latter in details here) due to interactions of electrons with various scattering potentials: interface roughness, alloy disorder, phonons, etc. as discussed in (5). It is worth to remark that if the gain cross-section gives the attenuation of the energy-density of the electromagnetic field due to the whole heterostructure as explained in [8], we have therefore to multiply $\alpha(\omega)$ by $N_p$, where $N_p$ is the number of periods and by $\Gamma$, the modal overlap as introduced in (8.3). The above expression for the absorption cross-section has the dimension of an inverse length.

The gain $g(\omega) = -\alpha(\omega)$ in this first-order model with homogeneous broadening is driven by the net population difference $\Delta n \equiv n_f - n_i$, as expected by the $k_\perp$-conservation. For two equally populated subbands the spectral gain is therefore zero for all frequencies.

It is worth to remark that since we have performed the rotating wave approximation, we have only extracted the pole given by: $\delta(E_{fi} - \hbar \omega)$. There indeed exist a conjugated pole given by: $\delta(E_{fi} + \hbar \omega)$. If $E_{fi} > 0$, the first pole corresponds to a positive frequency $\hbar \omega$ and the second to a negative frequency $-\hbar \omega$. Even if the energy of the free photon is positive: $\hbar \omega \geq 0$, the absorption is the sum of the two contributions. If only one contribution is considered, the zero-frequency absorption is ill-defined (the latter is however not well-described by a first-order model derived at optical frequencies).

Nevertheless if $E_{fi} \gg \gamma$, the influence of the negative pole is very weak and can be neglected if we evaluate the absorption in a few $\gamma$ around the resonant transition-energy. However if the relevant transition-energies (not necessarily the laser transition but also the absorption in the injector) becomes comparable to the broadening it is crucial to consider both poles.

This occurs typically in the design of far-infrared structures, but it is also relevant in the mid-infrared. As the free-carrier absorption, which is given by the inter-subband absorption in an heterostructure, represents the ultimate limit of quantum cascade lasers the value of the absorption in the injector-levels at the laser transition-energy has to be computed with accuracy.

8.2.2 Inhomogeneous broadening: application to first-order gain

We know that lifetime and dephasing give rise to a homogeneous broadening of the gain cross-section. The dirac line is replaced by a Lorentzian centered at the transition energy. This basic spectral shape is given by quantum mechanics for two-state systems with a finite
8.2. Gain mechanisms in heterostructures

lifetime. We have extended this results to a pair of subbands under some assumptions. In particular, we have considered that the heterostructure is a perfect repetition of a single period. We have therefore neglected all possible fluctuations in the width of the semiconductor layers. This long-range disorder however causes an additional broadening of the gain cross section \( g(\omega) = -\alpha(\omega) \) in Eq.(8.23). As we know that the gain-cross section cannot be resolved in space (for the z-direction) in the usual attenuation model [8], we have to implement fluctuations across the whole heterostructure on the gain cross-section directly. A simple model would be to assume that the resonant transition energy \( E_{fi} \) is subject to a random variation due to layer-thickness fluctuations. We have already modeled this problem for homogeneous broadening by computing the energy broadening due to interface-roughness, for example. However we assumed that this scattering potential was perturbing the quantum states and therefore caused transitions between them, yielding homogeneous broadening. This was discussed in the introduction of this section (8.2).

Here we consider long-range disorder where the transition energy fluctuate from period to period: we completely neglect the correlation between the upper and lower laser level between adjacent or more distant periods. This model is the simplest way to introduce inhomogeneous broadening. We therefore assume that the width of the wells is subject to a random distribution in a real heterostructure. Instead of computing the impact on the eigen-energies, we directly model a fluctuation of the transition energy \( E_{fi} \). If we assume that the transition energy for the period \( n \) is \( E_{fi}^{(n)} \), the gain cross-section will be averaged as:

\[
\langle g(\omega) \rangle = \frac{1}{N_p} \sum_{i=1}^{N_p} g^{(i)}(\omega),
\]

where \( g^{(n)}(\omega) \) is the gain cross section computed with transition-energy \( E_{fi}^{(n)} \). The average \( \langle g(\omega) \rangle \) is the incoherent superposition of the individual contributions. We usually prefer a continuous model for simplicity. The averaged cross-section is therefore given by a convolution between the distribution of transition energies and the homogeneous spectral gain.

We assume that the transitions energies are distributed following a normalized gaussian distribution:

\[
G(x; \mu, \sigma) \equiv \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right),
\]

where \( \sigma \) is the standard deviation (2\( \sigma \) representing the width of the gaussian distribution), and \( \mu \) the average. For the present problem we have \( \mu = \langle E_{fi} \rangle \), the averaged transition energy, and \( \sigma \) is given by: \( \sigma = \frac{1}{N_p} \sum_{i}^{N_p} (E_{fi}^{(i)} - \langle E_{fi} \rangle)^2 \). This parameter controls the
spectral broadening. The gain cross-section with inhomogeneous broadening is given by:

\[
\langle g(\omega) \rangle = \int_{E_0}^{\infty} dx \, g(\omega; x \equiv E_{fi}) G(x; \mu, \sigma)
\]  

(8.26)

Here, we have written the gain cross section: \( g = g(\omega; E_{fi}) \) as the average is made on the transition energy. We perform the integration on a semi-finite domain \([E_0, \infty] \): as discussed in the previous section, we cannot extend the integration to \(-\infty\) until we do not have considered the negative pole with the same inhomogeneous broadening. The energy \( E_0 \) can be considered as an infrared cut-off energy.

For numerical simulation, we rewrite the gain cross-section in Eq.(8.23) as:

\[
g(\omega) = \frac{G_0}{\hbar \omega} \frac{x^2 \gamma}{(x - \hbar \omega)^2 + \gamma^2}
\]

where

\[
G_0 = \frac{q_0^2 z^2 \gamma \Delta n}{L_p \varepsilon_0 n_r c \hbar}
\]

(8.27)

The peak-gain corresponds to \( \hbar \omega = x \): \( g_p = G_0 x/\gamma \). We impose that for \( x = \mu \) (where \( \mu \) is the averaged transition-energy) we have \( g_p = 1 \).

This allows to normalize the gain curves properly and to observe the reduction of the peak gain together with the variation of the spectral width and shape implied by inhomogeneous broadening.

In Fig.(8.1) we have illustrated the problem with a numerical example for a mid-infrared gain cross-section with an homogeneous broadening of \( \gamma = 6 \) meV and an (averaged)

![Figure 8.1: The effect of inhomogeneous broadening is illustrated for a mid-infrared transition energy. The details are given in the text.](#)
transition energy of $E_{fi} = 160$ meV. We have then varied the inhomogeneous broadening parameter from $\sigma = 0$ meV (no inhomogeneous broadening) to $\sigma = 4, 6, 10$ meV. For low values of $\sigma$ (with respect to the homogeneous broadening) the shape of the spectral gain is not modified but the peak-gain is reduced. For equal and higher values (here 6 meV and beyond) the shape is modified and the gaussian character overcomes the lorentzian shape. The peak-gain is by consequence largely reduced.

We expect the inhomogeneous broadening to be more important in far-infrared structures (the transitions energies and the homogeneous width being smaller) and may explain partially the discrepancy between the computed and measured linewidths.

8.2.3 Non-parabolicity effects on first-order gain

Until now we have neglected the effect of non-parabolicity on the shape of the spectral gain. The latter is introduced in the gain cross-section in Eq.(8.23), following [35], by considering the dependence of the transition energy in $k$. As $m_i \neq m_f$, we have:

$$E_{fi} = E_{fi}(k) = \varepsilon_f(k) - \varepsilon_i(k) = E_{fi}(0) + \frac{\hbar^2 k^2}{2} \left( \frac{1}{m_f} - \frac{1}{m_i} \right).$$

(8.28)

In order to implement this effect in the gain cross-section, we need to specify the intrasubband carrier-distributions in subbands $f$ and $i$. If we assume arbitrary distributions $f_i(k)$, we can write:

$$g(\omega) = \sum_k \frac{q_0^2 |z_{fi}|^2 E_{fi}(k)}{L_p \varepsilon_0 n_c \hbar \omega} \frac{\gamma [f_f(k) - f_i(k)]}{(E_{fi}(k) - \hbar \omega)^2 + \gamma^2}.$$  

(8.29)

The non-parabolicity modifies the transition-energy, however contrarily to the gaussian inhomogeneous broadening, this effect is triggered by the population difference $[f_f(k) - f_i(k)]$. If we assume thermal (Fermi-Dirac) distributions the spectral gain will not be altered by the non-parabolicity for low populations in subbands as the electrons mainly reside in states with low $k$-values. As the population increases non-parabolicity effects become stronger. The importance of the temperature has to be taken into account: for high temperatures, the spreading of the electron-distributions enable to populate states with high $k$-values.

The global effect of the non-parabolicity is a red-shift and an asymmetric broadening of the spectral gain. The peak-gain red-shifts as the upper-state mass is greater that the lower-state mass and therefore the detuning energy becomes smaller for higher $k$-values.
The broadening is asymmetric since the variation of the transition energy is not symmetric, contrarily to lifetime homogeneous broadening and to the type of inhomogeneous broadening considered in the previous section.

It is however important to remark that non-parabolicity is a homogeneous effect on the spectral gain as coherent processes can occur between subbands with different effective masses.

In order to clarify the situation, we consider that the upper subband has a net population $n_0$ and that the lower subband is empty. We consider an optical transition in the mid-infrared by choosing $E_{fi}(0) = 167$ meV and a homogeneous broadening of $\gamma = 6$ meV. The effective mass of the upper and lower states are $m_f \approx 0.058 m_0$ and $m_i \approx 0.048 m_0$ (these values were computed for a single-quantum-well active-region in the mid-infrared). Here we have kept the notation in usage for absorption processes as the final subband is the upper-subband.

We reparametrize the term under the sum in Eq. (8.29) as function of the final-state kinetic energy: $\epsilon_f(k) = \hbar^2 k^2/2m_f$. We have:

$$g(\omega) = \frac{G_{fi}}{\hbar \omega} \int_0^\infty d\epsilon_f \frac{E_{fi}(\epsilon_f) f(\epsilon_f)}{(E_{fi}(\epsilon_f) - \hbar \omega)^2 + \gamma^2}$$

where

$$G_{fi} = \frac{m_f q_0^2 |z_{fi}|^2 \gamma}{\pi \hbar^2 L_p \varepsilon_0 n_r c \hbar}.$$  (8.30)

The intra-subband distribution is given by: $f(\epsilon_f) = [1 + \exp(\beta(\epsilon_f - \mu_f))]^{-1}$. The fermi level $\mu_f$ is fixed by integration for a surface concentration of $n_0$ in the upper state. The detuning energy as a function of the kinetic energy in the upper subband is given by:

$$E_{fi}(\epsilon_f) = E_{fi}(0) + (1 - m_f/m_i) \epsilon_f.$$

For the above effective masses, the slope of the detuning energy is $\approx -0.21$. We compute the energy red-shift at the thermal energy $k_b T_e$ in the upper subband. For $T = 350$ K we have a red-shift of $-6.28$ meV from the transition energy at $k = 0$. We therefore expect a considerable effect.

If the temperature is high enough, we can approximate the Fermi-Dirac distribution by: $f(\epsilon) \approx e^{\beta\mu}e^{-\beta \epsilon}$. The term $e^{\beta \mu}$ can be taken out of the integral in Eq. (8.30): the shape of the spectral gain is therefore only determined by the detuning energy and does not depend on the net carrier concentration. We therefore consider the integral $s(\omega)$ given below. It gives the shape of the gain as a function of the ratio between the effective masses and as a function of the electronic temperature of the initial subband:

$$s(\omega) \equiv \frac{1}{\hbar \omega} \int_0^\infty d\epsilon_f \frac{E_{fi}^2(\epsilon_f) e^{-\beta \epsilon_f}}{(E_{fi}(\epsilon_f) - \hbar \omega)^2 + \gamma^2}.$$  (8.31)

In Fig. (8.2) we have computed $s(\omega)$ for various ratios of the effective masses: $m_f/m_i$. The red-shift and the asymmetric broadening of the gain shape are very clear.
Figure 8.2: The non-parabolicity effect on the gain curve is illustrated. The high-temperature limit was taken. For a fixed temperature of $T_e = 350$ K, the gain shape depends only on the ratio of the effective masses $m_f/m_i$, and on the thermal tail of the intra-subband carrier-distribution in the upper-level only. The effective mass ratio $m_f/m_i \approx 1.21$ corresponds to the single-quantum-well active-region discussed in the text.

We do not investigate the cases where is electron concentration are very high (for example $1 \cdot 10^{12}$ cm$^{-2}$). There, the red-shift and the broadening are very strong even at low temperatures. The ratio between the effective masses needs of course to be still appreciable. However these examples have no direct applications to quantum cascade lasers.

Another effect due to non-parabolicity is positive gain without a net population inversion. We have to mention that this effect should not be confused with second-order gain derived in the next section. The gain mechanisms are completely different, even if the final result is in both cases a local (in $k$-space) population inversion. The non-parabolicity gain relies on the variation of the local electron concentration as the subbands does not have the same curvature. However the in-plane momentum is still conserved in optical transitions. In second-order gain, local population inversion is obtained by an optical transition assisted by scattering. This enables a local population inversion in $k$-space as explained in the next section. However, in this case the in-plane momentum is not conserved.

In order to compute typical peak-gains obtained by pure non-parabolicity, we consider two
equally populated subbands (zero net population inversion). From Eq. (8.29) we have:

\[
g(\omega) = \frac{G_{fi}}{\hbar \omega} \int_0^\infty d\epsilon \frac{E_{fi}^2(\epsilon_f) [f_f(\epsilon_f) - f_i(\epsilon_i)]}{(E_{fi}(\epsilon_f) - \hbar \omega)^2 + \gamma^2}
\]

where

\[
G_{fi} = \frac{m_f \eta_0^2 |z_{fi}|^2 \gamma_{L_p}}{\pi \hbar^2 L_p \varepsilon_0 n_r c \hbar}.
\]

(8.32)

Where \( f_f(\epsilon_f) \) is the electron-concentration in the upper subband parametrized with the kinetic energy of the upper-subband \( \epsilon_f \). The distribution \( f_i(\epsilon_f) \) is the electron-concentration in the lower-subband parametrized by the kinetic energy of the upper-subband. It reads:

\[
f_i(\epsilon_f) = \frac{1}{1 + \exp \left( \beta \left( \frac{m_f \epsilon_f}{m_i} - \mu_i \right) \right)}.
\]

(8.33)

We have a net electron-concentration \( n_0 \) in both subbands. We also assume that we have a same electronic temperature \( T_e \). The fermi-levels are given by:

\[
\mu_\alpha = \frac{1}{\beta} \log \left( \exp \left( n_0 \frac{\beta \pi \hbar^2}{m_\alpha} \right) - 1 \right).
\]

(8.34)

The local population inversion as a function of the kinetic energy of the upper subband is given by: \( f_f(\epsilon_f) - f_i(\epsilon_i) \). If we integrate this difference in order to evaluate the effective population inversion \( \Delta n_{np} \) due to non-parabolicity, we get \( \Delta n_{np} \equiv 0 \). The net population inversion is zero.

In Fig.(8.3a) we have plotted the local electron-concentration as a function of the kinetic energy of the upper state, for a population of \( n_0 = 1 \cdot 10^{11} \) cm\(^{-2} \). The latter is a large concentration for usual mid-infrared quantum cascade lasers. We rather expect concentrations up to \( 5 \cdot 10^{10} \) cm\(^{-2} \). In Fig.(8.3b) we have plotted the corresponding gain curves for the parameters indicated in the caption of the figure. The gain due to non-parabolicity has a dispersive shape but is very weak in this system: the peak-gain is \( \approx 1 \) cm\(^{-1} \). As expected the peak-gain red-shifts and slightly increases with the temperature.

These calculations show that the dispersive gain observed in [95] cannot be explained by a first-order gain model with non-parabolicity.

### 8.2.4 Second-order gain

We discuss in this section the second-order gain model. It was investigated in a density-matrix model [115] and observed as a natural result in the non-equilibrium Green’s functions formalism [109]. A nice explanation is provided in [111] in terms of spectral functions. Although this explanation gives a very clear understanding of the origin of the gain between subbands, we prefer to discuss here its interpretation in the density-matrix formalism, following the guideline of this section.
Figure 8.3: The gain due to non-parabolicity between equally populated subbands \((n_0 = 1 \cdot 10^{11} \text{ cm}^{-2})\) in a single-quantum-well active region in the mid-infrared is illustrated for a range of electronic temperatures: 70 K, 150 K and 300 K. The single-quantum-well parameters are: \(m_f \approx 0.058 m_0, m_i \approx 0.048 m_0, E_{fi}(0) \approx 167 \text{ meV}, \gamma \approx 6 \text{ meV}, z_{fi} \approx 1.9 \text{ nm}, n_r \approx 3.4\) and \(L_p \approx 68.6 \text{ nm}\). (a) The local electron-concentration is plotted against the kinetic energy of the upper subband. (b) The corresponding gain curves are plotted.
We have seen previously that the first-order gain model is derived between a pair of sub-bands without other scattering potentials than the dipole-hamiltonian. The conservation of energy therefore imposes the conservation of the momentum in an optical transition. The second-order gain model is a natural generalization of this configuration as an additional elastic intra-subband scattering potential is considered. This is the same situation between the first and second-order currents (6.6). However the calculation of the second-order gain model in the density-matrix formalism is a bit cumbersome. We will only discuss the main steps and we left the reader with the original paper [115] for the details of the calculations.

The matrix elements of the hamiltonian between a pair of states $|n_k\rangle$ and $|m_k\rangle$ are found in Eq.(6.108).

$$\langle n_k'|H_0|m_k\rangle = \delta_{n,m}\delta_{k',k} \varepsilon_m(k) + \delta_{k',k} \hbar \Omega_{nm} + \delta_{n,m} \sum_q \delta_{k',k+q} V_q^m$$ (8.35)

The coupling with the electromagnetic field is introduced by the dipole approximation (8.1.1). The latter can be rewritten in a more general form as:

$$\delta h(t) = F_0 Q e^{-i\omega t},$$ (8.36)

where $F_0$ is the amplitude of the field, $Q$ is the operator that couples the electron system to the field and $e^{i\omega t}$ is the monochromatic time dependence. The operator $Q$ can be the position operator ($F_0$ is therefore the magnitude of the electric field) or the atomic momentum operator ($F_0$ is therefore the magnitude of the gauge potential). This hamiltonian is similar to a Peierls substitution. We consider that the current operator is obtained through the canonical relation:

$$J = -\frac{\delta H}{\delta F}.$$ (8.37)

If we identify $J = -Q$ and $J = e\dot{Z}$, this gives $Q = -\frac{ie}{\hbar}[H_0, Z]$. Such method enables to consider non-linear coupling between the electrons and the field. For example if we consider $H_c = \sum_{m=1}^{\infty} Q(m) F^m$, the coupling operator is determined with iterative commutation relations. The coupling operator at the mth order is given by:

$$Q_{(m)} = \frac{(-1)^m}{m!} \left(\frac{i e}{\hbar}\right)^m [[H_0, Z], Z], \ldots Z]_m \text{times.}$$ (8.38)

Since $[[H_0, Z], Z], \ldots Z]_m \text{times} = [Z^m, H_0]$, we can write:

$$H_c = [\exp\left(-\frac{ie}{\hbar} F Z\right), H_0],$$ (8.39)

which is similar to a Peierls substitution. Here we however retain only linear coupling with the field. We therefore use Eq.(8.36).
We compute the linear response with the complex field $e^{-i\omega t}$. However we have to consider at the end a real field by adding the adjoint operator $\delta h(t)\dagger$. The coupling Hamiltonian is therefore given by:

$$\delta H(t) = \frac{1}{2} (\delta h(t) + \delta h(t)\dagger).$$

(8.40)

Fortunately since everything is linear, we will only have to extract the real-part of the result. This is shown later in this section. For now, if we assume that the total Hamiltonian is given by:

$$\mathcal{H}(t) = H_0 + \delta h(t),$$

(8.41)

we can expand the density matrix in powers of the field amplitude: $\rho = \sum_{n=0}^{\infty} F_0^n \rho^{(n)}$. We insert this result in the equations of motion of the density-matrix operator. If we identify powers of the field, we have:

$$i\hbar \dot{\rho}^{(0)} = [H_0, \rho^{(0)}]$$

$$i\hbar \dot{\rho}^{(n)} = [H_0, \rho^{(n)}] + [Q, \rho^{(n-1)}]e^{-i\omega t}.$$  

(8.42)

The linear response is obtained by solving the equations for $\rho^{(0)}$ and $\rho^{(1)}$. The solution $\rho^{(0)}$ was obtained for the calculation of the current (6.6). For the linear part, we consider the Ansatz of the envelop operator: $\rho^{(1)}(t) = \lambda(t)e^{-i\omega t}$. It means that the system linearly responses at the same frequency than the probe field. The equations for the envelop function are therefore:

$$i\hbar \dot{\lambda} + \hbar \omega \lambda = [H_0, \lambda] + [Q, \rho^{(0)}].$$

(8.43)

We then directly apply the Laplace average (6.6.2) and we get the equations for an envelop-operator $\lambda$ in the Laplace representation:

$$i\hbar \lambda + \hbar \omega \lambda = [H_0, \lambda] + [Q, \rho^{(0)}].$$

(8.44)

The steady-state limit is obtained by the limit $s \to 0^+$ that gives: $\lambda(0^+)$. The solution for the first-order density-matrix is given by: $\langle \rho^{(1)} \rangle \approx \lambda(0^+)e^{-i\omega t}$. The solution for $\rho$ in steady-state is given by:

$$\rho \approx \rho^{(0)}(0^+) + \lambda(0^+)F_0 e^{-i\omega t} + O(F_0^2).$$

(8.45)

We then consider an observable $A$. Its expectation value is $\langle A \rangle_0$ when the field is off ($F_0 = 0$). The variation $\delta \langle A \rangle$ induced by the field is given by:

$$\delta \langle A \rangle(t) = \text{Tr} (\lambda(0^+)A) F_0 e^{i\omega t}.$$  

(8.46)

These variations are related to a transport coefficient, a susceptibility, as stated by the fluctuation-dissipation theorem. Here, if we consider that the coupling Hamiltonian has
the form $\mathbf{p} \cdot \mathbf{A}$, the electric field is related to the potential $f(t)$ by: $e(t) = -\partial_t f(t)$. We have $e(t) = i\omega F_0 e^{-i\omega t}$ and we can write the Ohm’s law in the Fourier representation by considering the variation of the expectation value of the current $J$:

$$\delta\langle J \rangle(t) = \sigma(\omega) e_0 e^{-i\omega t},$$

(8.47)

where $e_0 = i\omega F_0$ (i.e. $j_0 e^{-i\omega t} = \sigma(\omega)e_0 e^{-i\omega t}$) and $\sigma(\omega)$ is the complex conductivity.

We have considered a complex field in the hamiltonian. In order to recover the results obtained with a real field, we take the real part of the above equation (this gives the expected result since all equations are linear). We distinguish the real and imaginary part of the conductivity by writing $\sigma = \sigma_1 + i \sigma_2$. In the above equation this yields:

$$\Re[\delta\langle J \rangle(t)] = -\sigma_2 \omega F_0 \cos \omega t + \sigma_1 \omega F_0 \sin \omega t$$

(8.48)

It is worth to remark that the real-part of the conductivity (related to absorption) is in phase with the probe field while the imaginary-part (related to the refractive index of the medium) is out-of-phase. We finally have:

$$\sigma_1(\omega) = \frac{1}{\omega} \Im(\langle \lambda(0^+)J \rangle), \quad \sigma_2(\omega) = -\frac{1}{\omega} \Re(\langle \lambda(0^+)J \rangle)$$

(8.49)

We therefore have to solve the equations of motion for the envelop operator $\lambda$ in Eq.(8.44). The calculations are somewhat cumbersome but quite direct. They are similar to the calculations for the second-order current excepted that the photon energy appears in the diagonal part of the equations for $\lambda$. In [115] the author used a trick in order to extract the real-part of the conductivity. The same method was used in [55]. It relies on the fact that the coupling-hamiltonian has a similar structure as the non-diagonal part (tunneling) of $H_0$. The poles in the zero field polarization $f_{21}$ and $f_{12}$ are therefore shifted by $+\hbar \omega$ and $-\hbar \omega$.

We can define the absorption coefficient by the real-part of the conductivity:

$$\alpha(\omega) = \frac{\Re[\sigma(\omega)]}{\varepsilon_0 n_r c}.$$ 

(8.50)

This yields the spectral gain $g(\omega)$ for steady-state populations $f_{22}(\mathbf{k})$ of the upper-state (2) and $f_{11}(\mathbf{k})$ of the lower-state (1):

$$g(\omega) = \frac{\epsilon_0^2 d^2 |\Omega|^2}{\varepsilon_0 n_r c \omega} \sum_{\mathbf{k}} \frac{\gamma_1(\mathbf{k})(f_{22}(\mathbf{k}) - f_{11}(\mathbf{q}^+)) + \gamma_2(\mathbf{k})(f_{22}(\mathbf{q}^-) - f_{11}(\mathbf{k}))}{(\hbar \Delta - \hbar \omega)^2 + (\gamma_1(\mathbf{k}) + \gamma_2(\mathbf{k}))^2}.$$ 

(8.51)

Where $|\hbar \Omega|$ is the coupling (tunnel) energy between the upper subband and the lower subband and $\hbar \Delta$ is the detuning energy $\hbar \Delta \equiv \varepsilon_2(\mathbf{0}) - \varepsilon_1(\mathbf{0})$ evaluated at $\mathbf{k} = \mathbf{0}$. 


The broadening energies $\gamma^{(n)}(k)$ are defined by:

$$
\gamma^2(k) \equiv \pi \sum_q \delta(\varepsilon_2(q) - \varepsilon_1(k) - \hbar\omega)|V_{k-q}^2|^2 \\
\gamma^1(k) \equiv \pi \sum_q \delta(\varepsilon_2(k) - \varepsilon_1(q) + \hbar\omega)|V_{k-q}^1|^2.
$$

(8.52)

The special momentum $q_+$ and $q_-$ are defined by:

$$
q_+ = h^{-1}\sqrt{2m_1(\epsilon_2(k) + (\hbar\Delta - \hbar\omega))} \\
q_- = h^{-1}\sqrt{2m_2(\epsilon_1(k) - (\hbar\Delta - \hbar\omega))}.
$$

(8.53)

We have adapted the latter for subbands with different effective masses, with $m_2 \equiv m_f$ for the upper subband and $m_1 \equiv m_i$ for the lower subband. The in-plane kinetic energy is written: $\epsilon_a(k) = \hbar^2 k^2 / 2m_a$.

If we inspect the population terms in Eq.(8.51), we have:

$$
\gamma^1(k)(f_{22}(k) - f_{11}(q_+)) + \gamma^2(k)(f_{22}(q_-) - f_{11}(k)).
$$

(8.54)

The first term shows that population inversion is given between the electron concentration in the upper subband at wave-vector $k$ and the electron concentration in the lower subband

$$
\gamma^1(k)(f_{22} - f_{11}(q_+))
$$

Figure 8.4: The first population term of Eq.(8.51) is illustrated between a pair of subbands with non-parabolicity. The intra-subband scattering potential (inside $\gamma^1$) enables transitions from the upper subband (2) to the lower subband (1) out of the resonance condition: $\hbar\omega < \hbar\Delta(k)$. These transitions are allowed as the conservation of the momentum has been relaxed in optical transitions. In the figure the first-order transition is also represented in grey. The transition occurs at the detuning energy: $\hbar\omega = \hbar\Delta(k)$.
at wave-vector \( \mathbf{q}_+ \). The wave-vector \( \mathbf{q}_+ (\hbar \omega) \) depends on the photon energy at which we are probing the system. The situation is illustrated in Fig. (8.4) with non-parabolicity effects on the curvature of the subbands. The intra-subband scattering potential allows non-resonant optical transitions that conserve the energy:

\[
\varepsilon_1(0) + \frac{\hbar^2 k_{+}^2}{2m_1} = \varepsilon_2(0) + \frac{\hbar^2 k_2^2}{2m_2} - \hbar \omega. \tag{8.55}
\]

The second term in Eq. (8.54) has a similar structure.

This gain mechanism does not require non-parabolicity, although in real systems we have subbands with different masses. If we set \( m_1 = m_2 = m_w \), the second-order gain does not vanish.

We need to examine the structure of the spectral gain more precisely. For commodity we assume that the two subbands feel the same scattering potential by setting: \( \gamma \equiv \gamma_1 = \gamma_2 \). Rather than working in the tunneling basis (diagonal transitions), we consider vertical transitions with the substitution \( d^2/\hbar \Omega|^2 = |z|^2 (\hbar \Delta)^2 \), which is valid up to second-order terms in \( g(\omega) \). The gain cross-section is therefore given by:

\[
g(\omega) = \frac{g_0^2 |z|^2 (\hbar \Delta)^2}{\varepsilon_0 n_e \hbar^2 \omega} \sum_k \frac{\gamma(k)(f_{22}(k) - f_{11}(q_+)) + \gamma(k)(f_{22}(q_-) - f_{11}(q_+))}{(\hbar \Delta - \hbar \omega)^2 + (2\gamma(k))^2} \tag{8.56}
\]

The first-order gain is recovered formally by setting \( q_\pm = k \): the population difference is evaluated at the same wave-vector in both subbands.

The non-parabolicity plays an important role not in the building of a local (in \( k \)-space) population inversion, but in the shape of the spectral gain, as it was the case for first-order gain. It will be necessary to include the non-parabolicity effects on second-order gain when theoretical predictions will be compared to measurements.

If the following computations, we neglect the dependence of the broadening energy \( \gamma \) on \( k \). This approximation assumes that the majority of the electrons that contributes to the buildup of the spectral gain are in a sufficiently small portion of the reciprocal space where \( \gamma \) is essentially constant with respect to \( k \). The numerical value may however change with temperature. In the computation performed in this section we also neglect this variation \(^3\). The shape of the spectral gain is therefore not determined by a special \( \gamma \) function.

If we assume thermal intra-subband distributions, we can integrate directly the population term. For now we neglect the variation of \( \hbar \Delta(k) \) due to non-parabolicity, but we still keep different effective masses. The basic integral is given by:

\[
\int_{-\infty}^{\infty} dx \frac{1}{1 + \alpha e^{\beta(x-\mu)}} = \frac{1}{\beta} \log \left[ 1 + \frac{\alpha}{\beta} e^{\beta \mu} \right]. \tag{8.57}
\]

\(^3\) It is however taken into account in (8.4)
We have to integrate the terms $\sum_k f_{11}(q_+)$ and $\sum_k f_{22}(q_-)$. We define: $\delta \equiv \hbar \Delta(0) - \hbar \omega$. We have to ensure that $q_+^2 \geq 0$ and $q_-^2 \geq 0$. This yields:

$$\sum_k f_{11}(q_+) = \frac{m_2}{\pi \hbar^2} \left[ \theta(\delta) \int_0^\infty \frac{d\epsilon_2}{1 + e^{\beta(\epsilon_2 - \mu_1)}e^{\beta\delta}} + \theta(-\delta) \int_{-\infty}^0 \frac{d\epsilon_2}{1 + e^{\beta(\epsilon_2 - \mu_1)}e^{\beta\delta}} \right]$$

$$\sum_k f_{22}(q_-) = \frac{m_1}{\pi \hbar^2} \left[ \theta(\delta) \int_\delta^\infty \frac{d\epsilon_1}{1 + e^{\beta(\epsilon_1 - \mu_2)}e^{-\beta\delta}} + \theta(-\delta) \int_0^\infty \frac{d\epsilon_1}{1 + e^{\beta(\epsilon_1 - \mu_2)}e^{-\beta\delta}} \right]$$

(8.58)

We introduce the function $\chi(\delta) = \theta(-\delta) + \theta(\delta)e^{-\beta|\delta|}$, $\theta(x)$ being the Heaviside function, and we get for the population inversion $N(\hbar \omega)$:

$$2N(\delta) = \frac{1}{2} \sum_k \left[ f_{22}(k) - f_{11}(q_+) + f_{22}(q_-) - f_{11}(k) \right]$$

$$= \frac{m_2}{\pi \hbar^2} \frac{1}{\beta} \log \left[ \frac{1 + e^{\beta\mu_2}}{1 + \chi(\delta)e^{\beta\mu_1}} \right] + \frac{m_1}{\pi \hbar^2} \frac{1}{\beta} \log \left[ \frac{1 + \chi(-\delta)e^{\beta\mu_2}}{1 + e^{\beta\mu_1}} \right].$$

(8.59)

Due to the special definition of $\gamma$ in Eq.(8.56) compared to Eq.(8.23), we have to multiply the population term by a factor $\frac{1}{2}$ in order to get the population inversion defined as in first-order gain.

The population as a function of $\delta$ is plotted in Fig.(8.5) for parabolic approximation: $m_1 = m_2 = m_w \approx 0.042 m_0$. For equally populated subbands in Fig.(8.5a): the population inversion is zero at the resonance $\delta = 0$ and the first order solution is recovered: $q_\pm = k$. For positive $\delta$, i.e. lower photon energies, a population inversion is built-up and reaches: $+n_0/2$. We therefore expect gain on the lower energy side of the transition. The opposite configuration is found for higher photon energies where the population inversion reaches: $-n_0/2$. We therefore expect absorption on the high energy side of the transition. The gain is therefore dispersive shaped for equally populated subbands as shown in Fig.(8.7a)

In Fig.(8.5b) we have considered an empty lower subband. On the low energy side of the transition, we expect to recover the first order model. The maximal population inversion is reached and is equal to the net population in the upper subband $n_2$ as it is the case in the first order model. On the high energy side of the transition the population decays to the half of the net population of the upper subband: $n_2/2$. We therefore expect that the gain curve will drop faster with the photon energy on the high energy side of the transition. The sharpness of the decay is given by the thermal energy $\beta^{-1} \equiv k_B T_e$.

In Fig.(8.6), the same situation is considered but for $m_1 \approx 0.048 m_0$ and $m_2 \approx 0.058 m_0$, following the single-quantum-well example we have previously considered. In the case of equally populated subbands, a smaller population inversion is reached on the high energy side than on the low energy side of the transition. For empty lower subband, the half of the
Figure 8.5: The population inversion \( N(\delta) \) is plotted as a function of \( \delta = \hbar \Delta - \hbar \omega \) for various electronic temperatures: \( T_e = 50 \) K, 150 K and 350 K in the parabolic approximation \( m_1 = m_2 = m_w \approx 0.042 m_0 \) (a) For equally populated subbands \( n_2 = n_1 \equiv n = 1 \cdot 10^{10} \) cm\(^{-2}\). (b) For empty lower subband \( n_1 = 0 \) and populated upper subband \( n_2 = 1 \cdot 10^{10} \) cm\(^{-2}\). Details are given in the text.

The net population is still reached on the high-energy side of the transition, but a population inversion lower than \( n_2 \) is obtained on the low energy-side of the transition. The ratio is \( \approx 0.914 \) and does not depend on the electron density in the upper subband. This ratio is
Figure 8.6: The effect of non-parabolicity is considered for the configuration described in Fig. (8.5). The effective masses are $m_1 \approx 0.048 m_0$ and $m_2 \approx 0.058 m_0$. Details are given in the text.

related to the ratio between the effective masses. In a first-order model, the population inversion integrated on the kinetic energy in the subbands does not show this feature. The latter results from the evaluation of the intra-subband distributions at the special
wave-vectors $q_\pm$. For perfect population inversion we have:

$$N(\delta \geq 0) = \frac{1}{2} \sum_k \left[ f_{22}(k) + f_{22}(q_-) \right]$$

$$= n_2 \left( \frac{m_1 + m_2}{2m_2} \right).$$  \hspace{1cm} (8.60)

The discrepancy between the first- and second-order models appears clearly on the above equations.

In both situations (with/without non-parabolicity) the population inversion shows an increasing thermal spreading with temperature, as thermal intra-subband distributions are assumed. We therefore expect that on the one hand the dispersive gain (for equally populated subbands) will decrease with temperature and on the other hand that the first-order approximation will be recovered for an infinite electronic temperature. This gives the physical limit in which the first order gain is recovered.

Before we compute the gain directly it is worth to remind that in Eq.(8.56) we have only extracted the pole for positive frequencies: $\omega > 0$. The gain function however includes the negative pole, for $\omega < 0$. The latter is explicitly given in [115] and we will not reproduce it here for commodity. However in each computation we have taken care of the negative pole contribution when it was necessary. In fact the latter is relevant only when $\gamma \approx \hbar \Delta$ as we are forced to evaluate the gain near $\hbar \omega = 0$.

In Fig.(8.7) the second-order gain is evaluated for an optical transition at $\hbar \Delta = 167 \text{ meV}$ and a homogeneous broadening $\gamma = 6 \text{ meV}$. The period length of the heterostructure is $L_p = 68.6 \text{ nm}$, the dipole is $z \approx 1.9 \text{ nm}$ and the modal refractive index is $n_r = 3.4$. The non-parabolicity effects are neglected.

In Fig.(8.7a) the gain is computed for a zero net population inversion with equally populated subbands with an electron concentration of $n_0 = 1 \cdot 10^{10} \text{ cm}^{-2}$. The gain is dispersive shaped around the resonant-frequency as expected and the peak-gain decreases with increasing temperatures.

In Fig.(8.7b) the lower subband is empty and the upper subband has a net population of $n_2 = 1 \cdot 10^{10} \text{ cm}^{-2}$. The symmetry of the gain curve is progressively recovered with increasing temperature. The computed gain curves in Fig.(8.7c) are the most interesting. The transition between a dispersive gain with equally populated subbands (zero net population inversion) to an empty lower subband (pure net population inversion) is shown. The intermediate curves are taken at $n_1 = 0.8 \cdot 10^{10} \text{ cm}^{-2}$ and $n_1 = 0.5 \cdot 10^{10} \text{ cm}^{-2}$. The photon energy that corresponds to a zero gain corresponds to the transition energy for equally populated subbands, and then blue-shifts to the infinity when the population inversion is
Figure 8.7: The second-order gain is computed for a mid-infrared transition energy with a homogenous broadening of 6 meV. Non-parabolicity effects are neglected. The details are given in the text.

It is worth to compare the curves in Fig. (8.7a) to the ones obtained from pure non-parabolicity in Fig. (8.3) for a population ten times higher. For a same electron concentration the ratio between the peak-gains is about 80!

In the next sections we first examine the second-order gain with inhomogeneous broadening
and then non-parabolicity effects on both the populations and the transition energy.

8.2.5 Inhomogeneous broadening and second-order gain

We consider briefly the effect on second-order gain of inhomogeneous broadening with a gaussian model as in (8.2.2). In order to distinguish the various effects, we neglect non-parabolicity effects on the populations and on the transition energy. It appears clearly that for a strong population inversion, the effects are basically the same than for the first-order model. We therefore expect the gain curve to become more symmetric with increasing inhomogeneous broadening. The transition from the homogeneous line to the gaussian line occurs for a comparable or stronger inhomogeneous linewidth: \( \sigma \geq \gamma \). In these cases, even the sharp decay on the high energy side of the transition for the homogeneous line in the second-order model is lost in the averaging.

We expect a similar outcome when we consider the dispersive gain obtained for equally populated subbands in the second-order model. Indeed, if we multiply the gain curve with a symmetric distribution around the transition energy, we expect that the symmetric dispersive structure (we are neglecting here the non-parabolicity effects) is not affected. However, the transitions from homogenous to inhomogeneous (gaussian) character will also occurs for \( \sigma \geq \gamma \). We also expect that the peak-gain will red-shift and the peak-absorption will blue shift as we are summing dispersive functions. The peak-values are, in any way reduced.

However in intermediate cases, the photon energy that corresponds to a zero gain is slightly blue-shifted and does not correspond to the resonant transition energy even without additional broadening mechanisms. For these configurations we expect that the inhomogeneous broadening will alter considerably the spectral gain.

In order to give numbers, we have kept the example of the single-quantum-well mid-infrared transition considered previously with a homogenous linewidth of 6 meV. We have directly computed the spectral gain with inhomogeneous broadening. The results are shown in Fig.(8.8).

It is worth to remark that even with a symmetric inhomogeneous broadening the shape of the spectral gain is strongly modified and is in particular asymmetric. We therefore expect that non-parabolicity will yield even more complicated gain shapes.

The comparison of theoretical predictions with measurements therefore relies on the precise knowledge of the broadening mechanisms. However for mid-infrared structures we expect that the spectral distribution mainly results from homogeneous broadening with non-
8.2. Gain mechanisms in heterostructures

Figure 8.8: Inhomogeneous broadening (gaussian model) is applied on second-order gain. The values of sigma are reported on the figure. The value of the homogenous broadening is 6 meV. (a) When subband are equally populated. (b) Intermediate regime, the shape and the position of the peak gain are strongly altered. Details are given in the text. Non-parabolicity is neglected. It is worth to remark that the populations in the subbands are kept constant. We vary $\sigma$ only.

parabolicity and not from inhomogeneous broadening. However for sake of generality the latter should be also considered.
8.2.6 Non-parabolicity effects on second-order gain

In this section we examine the effect of non-parabolicity on second-order gain. The spectral distribution is modified in two ways: the population inversion \( N(\delta) \) in Eq.(8.59) depends on \( m_1 \) and \( m_2 \), and the transition energy between the two subbands depends on \( k \) as in Eq.(8.28):

\[
\hbar \Delta(k) = \hbar \Delta(0) + \frac{\hbar^2 k^2}{2} \left( \frac{1}{m_2} - \frac{1}{m_1} \right)
\]

(8.61)

where in the last line the transition energy was parametrized by the kinetic energy of the upper subband \( \epsilon_2 \).

The non-parabolicity also affect the population inversion as in Eq.(8.59). It is however worth to say that it does not modify the special wave-vectors \( q_{\pm} \). They are still defined as in Eq.(8.53) by:

\[
q_{+} = \hbar^{-1} \sqrt{2m_1 (\epsilon_2(k) + (\hbar \Delta(0) - \hbar \omega))}
q_{-} = \hbar^{-1} \sqrt{2m_2 (\epsilon_1(k) - (\hbar \Delta(0) - \hbar \omega))},
\]

(8.62)

since the conservation of the energy given in Eq.(8.55) has to be ensured.

The positive pole of the spectral gain therefore reads (normalized by the period length \( L_p \)):

\[
g(\omega) = \frac{1}{L_p \varepsilon_0 n_c c^2 \omega} \sum_k \frac{(\hbar \Delta(k))^{2\gamma}}{(\hbar \Delta(k) - \hbar \omega)^2 + (2\gamma)^2} \times \left[ f_{22}(k) - f_{11}(q_{+}) + f_{22}(q_{-}) - f_{11}(k) \right].
\]

(8.63)

where we have neglected the \( k \)-dependence of \( \gamma \). We consider thermal intra-subband distributions in both subbands with the same electronic temperature. In order to evaluate the spectral gain, we introduce the following parametrizations and definitions:

\[
K(\epsilon_2, \hbar \omega) \equiv \frac{(\hbar \Delta(\epsilon_2))^{2\gamma}}{(\hbar \Delta(\epsilon_2) - \hbar \omega)^2 + (2\gamma)^2} \quad \hbar \Delta(\epsilon_2) = \hbar \Delta(0) + \left( 1 - \frac{m_2}{m_1} \right) \epsilon_2
\]

(8.64)

\[
K(\epsilon_1, \hbar \omega) \equiv \frac{(\hbar \Delta(\epsilon_1))^{2\gamma}}{(\hbar \Delta(\epsilon_1) - \hbar \omega)^2 + (2\gamma)^2} \quad \hbar \Delta(\epsilon_1) = \hbar \Delta(0) + \left( \frac{m_1}{m_2} - 1 \right) \epsilon_1
\]

Following equations Eq.(8.58), we have to evaluate the following integrals, where: \( \delta \equiv \)
\[ h\Delta(0) - \hbar\omega. \] We define:

\[ I_2(\hbar\omega) \equiv \frac{m_2}{\pi\hbar^2} \left[ \int_0^\infty \frac{d\epsilon_2 K(\epsilon_2, \hbar\omega)}{1 + e^{\beta(\epsilon_2 - \mu_2)}} - \theta(\delta) \int_0^\infty \frac{d\epsilon_2 K(\epsilon_2, \hbar\omega)}{1 + e^{\beta(\epsilon_2 - \mu_1)}e^{-\beta\delta}} - \theta(-\delta) \int_{-\delta}^0 \frac{d\epsilon_2 K(\epsilon_2, \hbar\omega)}{1 + e^{\beta(\epsilon_2 - \mu_1)}e^{\beta\delta}} \right] \] (8.65)

\[ I_1(\hbar\omega) \equiv \frac{m_1}{\pi\hbar^2} \left[ \theta(\delta) \int_\delta^\infty \frac{d\epsilon_1 K(\epsilon_1, \hbar\omega)}{1 + e^{\beta(\epsilon_1 - \mu_2)}e^{-\beta\delta}} + \theta(-\delta) \int_0^\infty \frac{d\epsilon_1 K(\epsilon_1, \hbar\omega)}{1 + e^{\beta(\epsilon_1 - \mu_1)}e^{-\beta\delta}} - \int_0^\infty \frac{d\epsilon_1 K(\epsilon_1, \hbar\omega)}{1 + e^{\beta(\epsilon_1 - \mu_1)}} \right] \] (8.66)

Afterwards, the spectral gain reads:

\[ g(\omega) = \frac{1}{L_p} \frac{q_0^2 |z|^2}{\varepsilon_0 n_r c \hbar^2 \omega} \left( I_2(\hbar\omega) + I_1(\hbar\omega) \right). \] (8.67)

We have computed these integrals for the single-quantum-well active-region previously introduced. The results are shown in Fig.(8.9). The gain was computed for a ramping population in the lower subband, as explained in the caption of the figure. The results do not represent the population inversion in the measured structure [95]. We have postponed this discussion to the next section. Nevertheless by varying systematically the lower state population, we have a complete picture of both the second-order gain and of the effect of non-parabolicity on the second-order gain.

If we compare the curves, we observe an important red-shift of the peak-gain and of the peak-absorption. In the figure the bare transition energy is reported by a vertical line. With non-parabolicity effects, the gain curves are broadened and strongly asymmetric. The dynamic range of the peak-gain is reduced and the peak-absorption in increased. The red-shift appearing naturally in second-order gain curves with a lack of net population inversion is enhanced by the red-shift of the transition-energies due to non-parabolicity.

These computations show how the various parameters of the structure are important for obtaining theoretical gain curves closest to the measured ones.

In the next sections we report experimental evidences for second-order gain in mid-infrared quantum cascade lasers. We also establish the link between the Bloch oscillator and the second-order gain. The latter is named Bloch gain in the literature for this reason.

### 8.2.7 Second-order gain and the Bloch oscillator

In this section we establish the link between the second-order gain mechanisms and the usual semi-classical Bloch oscillator [10].
The non-parabolicity effects are computed for second-order gain. A mid-infrared transition at 167 meV with a homogeneous broadening of 6 meV is considered. The dipole is $z = 1.9$ nm, the period length $L_p = 68.6$ nm. The modal index is $n_r = 3.4$. The gain curves were computed for an upper state population of $n_0 = 1 \cdot 10^{10}$ cm$^{-2}$. The population of the lower state is then varied as a fraction $x$ of $n_0$. The curves are displayed between $x = 0$ and $x = 1.0$ with a step of 0.1. (a) The non-parabolicity effects are neglected: both subbands have a mass $m_w$. (b) Non-parabolicity effects are considered on the populations and the transition energies. The results are discussed in the text.
In solids electrons have a fixed relation between momentum and energy: they are moving along energy bands as known from condensed matter theory. When an electric field is applied they are accelerated but the lattice forces a periodic motion at a definite Bloch frequency. This phenomenon is known as Bloch oscillations, and the idea was successfully used by Zener to explain the dielectric breakdown [121]. However in usual solids the strong scattering due to impurities and carrier-carrier interaction prevents the observation of such oscillations, as the lattice constant is too short to allow the electrons to complete even one oscillation cycle. In superlattices, the lattice constant can be chosen and a subtle engineering may allow electrons to achieve a few oscillations before scattering. As this phenomenon is fascinating from a condensed matter point of view, it also opens new perspectives for optics since charge oscillations naturally couple to radiation and offer a way to emit coherent radiation.

The important question is therefore whether these oscillations can be self-sustained and provide optical gain. First Ktitorov [60] and then Ignatov and Romanov [43] addressed the problem theoretically with Boltzmann equations and succeeded to provide a definitive signature for Bloch oscillations in superlattices in terms of a particular spectral response: the Bloch oscillations are found to amplify the electromagnetic field (optical gain) on the low energy side of the oscillation frequency, while they absorb photons on the high energy side. This particular shaped gain - Bloch gain - is the main feature, of the Bloch oscillator. A series of experiments [32, 85, 114] using pulsed ultrafast techniques have successfully shown the existence of Bloch oscillations as electrons are pumped in a higher energy band and collectively oscillate over their dephasing time. However the Bloch gain extends to zero frequencies and the structure becomes unstable in steady-state, so far preventing the observation of net gain in superlattices, although some evidence in photocurrent [102] and more recently in absorption [82] were demonstrated by driving superlattices with a strong terahertz field.

Recently Bloch gain was investigated in a pure quantum mechanical perspective in a density-matrix formalism, as reported by Willenberg et al. [115], and also in a Green function formalism, as reported by Wacker [109]. These theoretical works demonstrated that Bloch gain can arise between any pair of states, called subbands, in a semiconductor heterostructure and that the latter does not need to be a superlattice. As shown in Fig. (8.10a), the origin of this gain are second-order, scattering-assisted, optical transitions occurring between pairs of subbands. In an emission process, electrons at wavevector $k$ in the upper subband perform vertical optical transition to a virtual lower subband, and then scatter elastically to a real state at wavevector $k'$. The absorption process proceeds in a symmetric manner, exchanging the upper and lower subbands Fig. (8.10b). The essential feature of these second-order processes is that they connect now states with different
Figure 8.10: Bloch gain mechanism (a), (b): Scattering-assisted optical transitions between a pair of subbands with conduction band non-parabolicity. Black parabola represent real energy states, pink parabola are virtual states dressed by scattering.

wavevectors. As a consequence, as compared to direct transitions, the population-inversion requirement for these second-order processes are relaxed, since population inversion is only necessary between states with different wavevector and not globally between subbands. For example, in the particular case were both subbands have the same populations, gain is achieved for photon energies smaller than the transition energy as the occupation at the final state $k'$ is automatically smaller than the one of the initial state $k$. Note that because of the short upper state lifetime in quantum cascade lasers, populations and occupations at typical electron temperatures are small, and therefore Pauli blocking is negligible although it is implicitly contained in the theoretical models. As these theories do not require any specific shape for the heterostructure potential, Bloch gain may also be observed in a sample where an injection region have been inserted between active wells in order to stabilise electrically the structure, as it is routinely done in quantum cascade lasers [30]. Moreover the model could treat subbands with arbitrary populations, from perfect inversion (no lower state population), equal population (in which case it was shown numerically to yield the same results as Boltzmann models), to absorption (no upper state population) and therefore this second-order picture generalises the concept of gain in semiconductor heterostructures.

The gain mechanism is therefore the same in a traditional Bloch oscillator in a superlattice than between subbands in a quantum cascade laser. This has motivated to named this gain mechanism bloch oscillator gain or simply bloch gain. We therefore have designed
special quantum cascade structures that enable the population inversion to be varied over a large range. We therefore expect to reproduce experimentally all the regimes shown in Fig. (8.9).

8.2.8 Bloch gain in a mid-infrared quantum cascade laser

To demonstrate experimentally the existence of Bloch gain in quantum cascade laser structures, and at the same exclude other phenomena that were predicted to also yield gain without global population inversion between subbands, such as non-reciprocity [44] and local k-space inversion due to non-parabolicity [28, 37], a spectrally resolved gain measurement is performed in a specially designed quantum cascade laser. Based on a comparison of laser characteristics, laser operation without global population inversion between subband was already reported [28] and was attributed to the combined effect of non-parabolicity and a hot electron distribution in the lower state [37]. However, no measurement of the gain curve were performed in this experiment, preventing a definitive interpretation of these results.

In this work, in contrast to quantum cascade lasers that are usually optimized to ensure a strong population inversion, our structure was designed to enable the population inversion to be varied over a large range. To this end, as shown in Fig. (8.11), the active period consists of an injection region coupled to an active quantum well by an injection and extraction tunnel barrier (the parameters of this active region were used previously to compute first-order and second-order gain with various broadening sources). Under the designed bias of 48 kV/cm, the band-structure is aligned so that electrons tunnel resonantly from injector ground state $|g\rangle$ of the previous period to upper lasing state $|2\rangle$. As shown in [88], the thickness of the injection barrier must be optimized: in the low coupling regime (thick injection barrier), the maximum current that can be injected in the upper state is proportional to the coupling strength; in the opposite (thin injection barrier) the maximum current is only limited by the upper state lifetime but leakage currents are then proportional to the coupling. Downfield of the active well, an extraction barrier allows to control population of the lower lasing state $|1\rangle$: since $|1\rangle$ is the ground state of the active well, coupling to the upper state $|u\rangle$ of the injector region is achieved by tunneling of electrons through the barrier; by varying the width thereof, the extraction time and therefore the population of the lower state can be tuned. A submonolayer, 0.2 nm thick Al$_{0.48}$In$_{0.52}$As barrier, inserted in the active quantum well, raises the energy of the lower state by about 18 meV, so that it faces the extraction miniband. The presence of this barrier does also enhance the elastic broadening of the lower state.
Figure 8.11: Sample details. Layer sequences and band-structure of the investigated sample. The structure is shown at injection resonance field. The layer sequence starts from the injection barrier and the thicknesses are in nm; roman, resp. bold, numbers indicate In\textsubscript{0.53}Ga\textsubscript{0.47}As, resp. Al\textsubscript{0.48}In\textsubscript{0.52}As alloy, acting as well, resp. barrier material. Underlined layers are $3 \times 10^{17}$ cm$^{-3}$ Si doped. N258 single quantum well structure with layers, from right to left: 4.8/3.6/0.2/3.6/3.5/5.1/1.1/5.0/1.2/4.5/1.3/3.5/1.5/3.4/1.6/4.4/1.8/3.2/2.1/3.0/2.5/3.0/2.9/2.9. Nominal sheet carrier density is $3.03 \times 10^{11}$ cm$^{-2}$. Period length is 68.6 nm, repeated 35 times.

A systematic study of injection and extraction barrier widths has been carried out. It allows to correlate the characteristics of the device, as well as the shape of the gain curve, with the lower state lifetime. As expected, a reduction of the population inversion is observed when extraction-barrier width is increased, translating into an increase of the laser threshold. Here, we choose to focus our study on sample N258 that achieved lowest population inversion while still showing laser action. As a reference, we also studied a sample based on a two-well active region where a strong population inversion is expected [90]. As shown in Fig.(8.12), the design of sample N123 relies on a strong depopulation of the lower lasing state $|2\rangle$ through an intersubband LO-phonon resonance (32 meV) between $|2\rangle$ and $|1\rangle$ which keeps the lasing doublet $|3\rangle$, $|2\rangle$ in a strong population-inversion regime.

The samples were grown using the Ga\textsubscript{0.47}In\textsubscript{0.53}As /Al\textsubscript{0.48}In\textsubscript{0.52}As material system lattice matched on InP substrate and consisted in 35 periods of the injector/active region pair placed in the center of an optical waveguide. As indicated in the caption of Fig.(8.11, 8.12), a few layers in the center of the injectors were Si-doped in order to yield a sheet
8.2. Gain mechanisms in heterostructures

Figure 8.12: Sample details. Layer sequences and band-structure of the investigated sample. The structure is shown at injection resonance field. The layer sequence starts from the injection barrier and the thicknesses are in nm; roman, resp. bold, numbers indicate In$_{0.53}$Ga$_{0.47}$As, resp. Al$_{0.48}$In$_{0.52}$As alloy, acting as well, resp. barrier material. Underlined layers are $3 \times 10^{17}$ cm$^{-3}$ Si doped. N123 two-well structure with layers, from right to left: 4.5/8.0/1.0/5.7/2.4/4.4/1.4/3.6/1.2/3.6/1.2/3.4/1.0/3.4. Nominal sheet carrier density is $2.82 \times 10^{11}$ cm$^{-2}$. Period length is 48.4 nm, repeated 35 times.

density of $n = 3 \cdot 10^{11}$ cm$^{-2}$ per period. Samples were then processed into laser ridges using wet etching, Si$_3$N$_4$ insulation and standard processing techniques.

The gain measurements were performed using a variant [38] of a multi-section cavity technique successfully applied on GaAs based QC lasers [6]. Samples are processed in four-section ridges: an absorbing section, a light bulb section (A), an insulating (or middle) section (M) and an amplifier section (B). To decrease the electrical cross-talk between sections, the heavily doped cladding region was removed over a width of about 20 µm by a dry etching process. During the whole measurement, the absorbing section was grounded to damp optical feedback and prevent the device from oscillating. The section A, used as a light bulb, was driven above the onset of negative differential resistance (NDR). In this manner, the injection process is non-resonant and does not only populate the upper state but also the quasi-continuum of states lying above, providing a relatively broad emission spectrum. The middle section is biased to an intermediate bias where no significant current is injected, but also where the ground state of the active region, lifted above the Fermi energy of the injector, is essentially empty, preventing reabsorption. Electroluminescence measurements are performed in step-scan using a Nicolet 860 Fourier-transform
infrared spectrometer. A selected sequence of pulses [38], combined with a phase-sensitive detection, enables the simultaneous measurement of the electroluminescence spectrum of section A amplified by section B, minus the direct electroluminescence of section B, and the emission of section A alone. The net modal gain of the amplifier section can then directly be extracted from this measurement as explained in [38].

The measurements, performed at liquid helium temperature, are shown for the single quantum well sample (N258) as a function of injected current in Fig.(8.13a). At the lower current densities (2.1 and 3.3 kA/cm$^2$), the gain shows a clear dispersive shape, with gain for photon energies below the expected transition energy and loss at higher energies. As the current is increased, a more symmetric shape is recovered, indicating an improvement in the population inversion as the injection is made more resonant.

Figure 8.13: Net modal gain measurements for N258 and N123 samples (a) (N258) Measurements achieved for increasing current densities in the amplifier section: 2.1, 3.3, 4.5, 5.7, 6.9 kA/cm$^2$. The dashed vertical line represents the transition energy computed from the band-structure between subband edges ($k = 0$ states) of the active well. Shown also in (a) is the spectrum of device N258 operated as a laser. (b) (N123) Measurements achieved for increasing current densities in the amplifier section: 2.1, 3.0, 4.0, 5.5, 7.8 kA/cm$^2$. The simulations published in [95] are also shown in the bottom panels.
Gain measurements of the control sample N123, are shown in Fig.(8.13b). As expected, the shape of the gain curve is almost symmetric for the whole range of injected current. For a given injection current density, the maximum gain is also larger by a factor of 4.

The dispersive shape of the gain spectrum, and its dependence on the active region design, are a clear demonstration of the presence of Bloch gain in the low inversion sample. An absorption artefact that could simulate the dispersive shape shown by sample N258 would be measured equally in both samples if it was process or impurity-related. In addition, careful computation of the absorption in the injector region for varying temperature and biases, may safely exclude the presence of a parasitic absorption line at such photon energy. As an additional check, the device N258 was then cleaved and operated as a laser; as expected its spectrum is centered at the photon energy at which the maximum of the gain was recorded, as shown in Fig.(8.13a).

The validity of our interpretation is further verified by computing the gain theoretically. The result of this computation is shown in Fig.(8.13) for sample N258 and for the reference sample N123. For the comparison to be meaningful, a procedure that used essentially no fitting parameters was implemented. The shape of experimental and theoretical curves are found in a very good correspondence. In comparison, the prediction for sample N123 shows a line that is much more symmetrical and closer to a Lorenzian shape, again mirroring the experimental results. Although, for both samples, the gain shapes are in a good agreement with the experimental data, the absolute value of the gain is overestimated by a factor of 3 in each case. We attribute a part of this discrepancy to the experimental technique used to measure the gain spectra: parasitic light, emitted by the light bulb section A, travelling through the substrate and collected by the lens will increase the intensity of the measured light for the normalization, in effect leading to an underestimation of the gain. Experimental evidence for this effect is obtained by comparing the waveguide losses measured with the multisection technique (5 cm$^{-1}$) with a systematic measurement of the threshold current as a function of cavity length (11 cm$^{-1}$). The latter value is also in good agreement with the computed one using a Drude model for free carrier absorption. On the theoretical model side, neglecting all other scattering processes except for optical phonon scattering leads to an overestimate of the gain.

The magnitude of the gain measured in this work is much larger than the one predicted for non-parabolicity only as shown in Fig.(8.3b) and in [37] and the shape of the gain curve is clearly different. In our model, although inclusion of the non-parabolicity was important to achieve a good correspondence between the shape of the experimental data and the theory, the key ingredient needed to predict both dispersive shape and right magnitude of the gain curve is the Bloch component of the second-order gain and not the
Lasers were also fabricated from sample N258 and N123 and their optical and electrical characteristics measured as a function of temperature and shown in Fig.(8.14). The threshold current density of sample N258, at low temperature, is 2.2 kA/cm² for a 3.8 mm long and 26 µm wide device. Even allowing some uncertainty on the current values displayed in Fig.(8.13) because of current spreading, the device operates clearly in the regime of Bloch gain, as inspection of the gain curve measured at $J = 3.3$ kA/cm² shows a strong dispersive component. The device operates up to a maximum temperature larger than 300 K, with a weak temperature dependence of the threshold current density that fits better a linear rather than the usual exponential dependence $J = J_0 \exp(\frac{T}{T_0})$. Fitting the data nevertheless for $T > 240$ K with this expression yields a $T_0 = 265$ K. Sample N123, in contrast, displays the usual exponential dependence of the threshold current density with a value of $T_0 = 107$ K, as also reported in [90]. Although more detailed temperature dependent measurements and modeling would be required to rigorously prove that the temperature behavior is indeed related to the nature of the gain [39], a weaker temperature dependence of the threshold current density is qualitatively expected from a device based on a Bloch gain mechanism. The threshold current density at 300 K of N258 is a factor of about five than recent results achieved in optimised structures, yet the level of performance of sample N258 is still suprising as its doping level and injection-barrier thickness were optimized for the visualization of Bloch gain and not for sheer performance.

Figure 8.14: (a) Bias (left axis) and peak optical power (right axis) as a function of temperature for the single quantum well N258. (b) Threshold current density as a function of temperature, for both samples N258 (3.8 mm long) and reference sample N123 (2.6 mm long). The segments show the temperature range over which the $T_0$ parameter was fitted.
The experimental evidences for Bloch gain are very clear, however the simulations performed at the time of writing [95] were too rough to match the experimental curves even by including non-parabolicity effects. Nevertheless the shape of the gain curves and the dynamic range (if scaled to the experimental one) were well-predicted for both samples. The following paragraphs describe the method used for the simulations.

The gain computations were performed, in a first step, by computing the band-structure self-consistently assuming thermal populations. In a second step, the electron population is computed for each level through rate equations, assuming for both electron injection and extraction a resonant tunneling formalism based on density matrices [56, 88] (6.5). In this computation, the population lifetimes were computed assuming bulk optical phonon emission. In-plane scattering times were taken from the measured width of the electroluminescence: 0.13 ps (10 meV) for sample N258 and 0.2 ps (8 meV) in sample N123. The influence of the populations on the upper and lower laser states on the band-alignment was neglected, i.e. only the thermal populations were taken into account. Once the populations were obtained, the second order gain was computed including the non-parabolicity effects. The second order gain at frequency $\omega$ can be written as [115]:

$$g(\omega) = \frac{e^2 |z_{21}|^2 \epsilon^2}{L_\rho \varepsilon_0 n_r c \hbar^2 \omega} \sum_k \frac{\gamma_k^1 (f_{k+}^{22} - f_{k+}^{11}) + \gamma_k^2 (f_{k-}^{22} - f_{k-}^{11})}{(\epsilon - \hbar \omega)^2 + (\gamma_k^1 + \gamma_k^2)^2},$$

where $z_{21}$ is the dipole-matrix element, $\epsilon$ the intersubband transition energy, $f_{k+}^{ii}$ the distribution function of state $i$ at wavevector $k$, $\gamma_k^i$ the homogeneous broadening of state $i$ at wavevector $k$; $e$ is the electron charge, $\varepsilon_0$ the vacuum permittivity, $n_r$ the refractive index, $c$ the speed of light, $\hbar$ Plank’s constant divided by $2\pi$. The gain expression consists in two terms: an emission term $f_{k+}^{22} - f_{k+}^{11}$ and an absorption term $f_{k-}^{22} - f_{k-}^{11}$, where the scattering assisted transitions appear through the special wavevectors $k_{\pm}$ depicted in Fig.(8.10). In order to include the non-parabolicity, all energies, including in the wavevectors $k_+$ and $k_-$ at which the distribution functions are evaluated, are now computed assuming an energy dependent effective mass with the form $m^*(E) = m^*(0)(1 + E/E_G)$ where the $E_G = 0.8$ eV is the energy gap.

### 8.2.9 Refined transport simulations for N258 sample

In this section we apply the transport model developed in (7) on N258 sample investigated in the previous section for bloch gain. First of all we try to reproduce the IV curves shown in Fig.(8.14) for a wide range of lattice temperatures. In the simulation we consider resonant tunneling at the injection and at the extraction barrier as shown in Fig.(8.11).
A uniform electronic temperature is considered for all subbands and compute with the kinetic energy balance model developed in (7.2). For scattering mechanisms, we consider interface roughness with parameters $\Lambda = 9.0$ nm, $\Delta = 1.2$ Å, and $\kappa = 1.5$ nm, alloy disorder, ionized impurities scattering and LO-phonon with an energy of $\hbar \omega_{\text{LO}} = 32$ meV (5). Practically, we have implemented the transport model on a cluster$^4$ which enables to solve simultaneously the transport for each electric fields and each temperatures. This has no impact on the results but as we will discuss in (8.5) this has a strong impact on the feasibility of an automated optimization of structures. For one point our simulator takes about minutes on modern hardware$^5$.

The current-density versus applied electric field curves ($J - F$ curves) are shown in Fig.(8.15) together with the electronic temperature as a function of the current density and the lattice temperature. The population in the upper and lower laser states are also reported.

The simulated current density is very close to the experimental one over a large range of temperature, from 15 K to 300 K. We can observe the same crossing point near 7 kA/cm$^2$. The latter corresponds roughly to the extraction disalignment. Regarding the electronic temperature, the latter shows large variations with the current density. Even at a temperature of 15 K, $T_e$ reaches a value of 300 K at 7 kA/cm$^2$. It also plays an important role in the agreement between simulated and experimental currents. The $J - F$ curves computed without the self-consistent computation of the electronic temperature cannot reach a current of 6 kA/cm$^2$ and the crossing point is not observed. It is worth to recall that we have no adjustable parameters here and we cannot therefore try to bring the simulations closer to the experiment. In particular, as shown in Fig.(8.16), the simulations are largely pessimistic compared to the measured gain curves in Fig.(8.13). We probably underestimate the efficiency of the extraction mechanism, or the upper-state lifetime. The modeling of the spike in the middle of the active-well is possibly not accurate enough. Here we do not want to play with the vertical correlation as it was computed correctly (based on a physical argument) in (5.3.1). However the differential gain was shown to depend weakly on the temperature [39]. Even if the absolute value of the populations is not as good as expected, we however observe that the population inversion does not depend strongly on the temperature for high temperatures ($\geq 200$ K).

On the one hand, we have largely overestimated the peak gain with a simple model as shown in Fig.(8.13), although the measured peak gain was somewhat underestimated. And on the other hand, we are underestimating it with a more sophisticated model, that

---

$^4$ETH Brutus cluster: brutus.ethz.ch

$^5$Intel core-duo first generation (2006)
8.2. Gain mechanisms in heterostructures

Figure 8.15: Transport simulations of sample N258 are shown for lattice temperatures: 15, 80, 120, 160, 200, 240, 300 K. (a) The applied electric field is shown as a function of the current density. (b) Corresponding experimental measurements. (c) The upper/lower state net population $n_2/n_1$ are shown as a function of current density. (d) The uniform simulated electronic temperature is shown as a function of current density. The color code from measurements in (b) is used for all panels.

however predict correctly the current-density in the structure. It is difficult to provide a complete explanation (without performing more simulations and measurements), but it appears quite clearly that we cannot ascribe thermal populations to the active subbands. We need an electronic distribution resolved in the reciprocal space together with bloch gain mechanism to explain satisfactorily sample N258. We have tried to attribute a different temperature to the upper and lower subband and impose a zero kinetic energy balance on each subband. The results were meaningful, the temperature of the lower subband diverges: no solutions can be found for temperatures below 1500 K. This is a direct evidence of the existence of a non-thermal intra-subband distribution in the lower subband.
Figure 8.16: The bloch gain is computed for populations found in the transport simulations (8.2.9) at 15 K. The current densities are: 1.7, 3.2, 5.1, 6.1 and 7.1 kA/cm².

Anyway, we evaluate the bloch gain curves for the populations of the upper and lower active subbands and the electronic temperatures obtained in the simulation for a lattice temperature of 15 K. For consistency with the previous simulations a homogeneous broadening of $\gamma_6 = \text{meV}$ is used, in agreement with the experimental curves and the computed value of 6.5 meV shown in Fig. (5.2). The results are shown in Fig. (8.16). If we compare them to measurements in Fig. (8.13), we observe, as discussed above, a large discrepancy. Moreover, the bloch gain near zero population inversion is very sensitive to many parameters of the model. The simple four-state model developed for the far-infrared in (6.5) gives better results, as the extraction process from the lower subband is very simplified and is therefore approximated with effective parameters. This yields an overestimated extraction efficiency.

We have invoked previously the thermal intra-subband distributions in active subbands as a limitation of our transport model. This is of course right. However the discrepancy, in particular the too low population inversion can be partially explained by incriminating the coupling parameters are the extraction barrier. We may assume that the tight-binding model (4.4) is not accurate enough in this case. This can be due to the first-order approximation or to the fact that the extraction cannot be modeled properly by considering only pair of states. Disregarding the model limitations, the problem can even be related to a
growth issue. This is less plausible. In any cases, we may consider that the extraction barrier has an *effective width* which is typically thinner than the nominal one.

We have therefore simulated the sample N257. It has the same layer sequence than N258 excepted for the extraction barrier. The latter is only 3.0 nm instead of 3.5 nm for N258. The results are shown in Fig.(8.17). The $J - F$ curves are very similar to the ones of N258 sample. This is expected as the injection barrier was not modified. They also show a crossing region, corresponding roughly to the extraction disalignment. The electronic temperature curves are not shown as they are basically the same for both samples. Regarding the populations in the active subbands the situation is completely different as a net population inversion is observed up to 240 K. This is what we expected: the extraction mechanism has a high sensitivity and the configuration in the real sample is not known with enough precision to model the populations of the active subbands with the required accuracy, in order to match measurements and simulations. The problem regarding the thermal intra-subband distributions is of course unsolved.

We have computed the gain in sample N257 for the populations and (electronic) temperatures obtained at 15 K. The results are shown in Fig.(8.18). They are closer to measurements, as we observe clearly the transition between absorption, dispersive gain and gain based on a strong population inversion.

If we consider the results for N258 and N257 in Fig.(8.18) we observe that the linewidth of the gain curve is overestimated and that the peak-gain is strongly red-shifted. This is

![Figure 8.17](image-url)

Figure 8.17: Transport simulations of sample N257 are shown for lattice temperatures: 15, 80, 120, 160, 200, 240, 300 K. (a) The applied electric field is shown as a function of the current density. (b) The upper/lower state net population $n_2/n_1$ are shown as a function of current density.
Figure 8.18: The bloch gain is computed for populations found in the transport simulations (8.2.9) at 15 K. The current densities are: 1.9, 3.5, 5.2, 6.1 and 7.2 kA/cm².

not observed in measurements in Fig.(8.13). This effect is clearly ascribed to the overflow of the electronic temperature in the active subbands: as electrons populates artificially states with a larger $k$ vector, the non-parabolicity effects are overestimated.

The conclusion of this discussion is that we were able to found an experimental signature of the bloch gain in a samples based on single-quantum-well active region designed to emit in the mid-infrared.

On the theoretical side the necessity of second-order mechanism for describing the gain and the absorption (equivalent to free-carrier absorption in bulk materials) is crucial. Moreover this clarify the quantum mechanical mechanisms on which the gain in the Bloch oscillator relies.

We have modeled the $J − F$ curves very accurately for a large range of temperatures. However the extraction mechanism is subtle and we had to modify the width of the extraction barrier (by 0.5 nm) to obtain gain curves much closer to measurements. We have shown that the non-parabolicity affects strongly the shape and the peak-value of the spectral gain. We have shown that modeling the transport with thermal intra-subband distributions leads to an overestimation of the electronic temperature. This is an evidence of the out-of-equilibrium nature of the intra-subband carrier-distributions in active subbands.
8.3 Computation of light-current curves in the far-infrared

In this section, we implement the coupling between the laser field in an optical cavity and the electrons. Provided we are able to compute the spectral gain in transport configurations, we will be enable to compute the threshold current density and the optical power emitted by the structure. However these computations require some knowledge on the laser cavity, as the modal overlap $\Gamma$, the modal index $n_r$ and the total optical losses $\alpha_{\text{tot}}$, due to the optical cavity and possibly to the active region. We develop here a minimal model which is well-adapted to ridge-waveguide configurations. Nevertheless it may be used conceptually for treating any situation were the optical mode is stable.

We first consider the case of a far-infrared structures based on resonant tunneling. We have already modeled the transport part in (6.5). A special feature (double threshold currents) in the light characteristic will enable the determination of relevant lifetimes in the structures by fitting the model to measurements.

Before we give more details, we consider very briefly the characteristic lifetime of a state due to spontaneous emission. We have discussed its origin and we have use it in order to obtain the linewidth due to lifetime broadening in (8.2.1). The spontaneous emission is the initiator of the laser oscillation. However in quantum cascade lasers the spontaneous emission lifetime is very long compared to other time scales. Even if the computation of the emitted power requires in principle to solve the time evolution of the optical intensity in the cavity, we can consider only absorption, stimulated emission and a steady-state photon flux density as shown below.

The spontaneous emission rate for two states separated by energy $E_{f_i}$ and optically coupled by the oscillator strength $f_{ij}$ is given by:

$$W_{sp} = \frac{q_0^2}{6\pi m_0 c^3 \varepsilon_0 \hbar^2} n_r f_{ij} E_{f_i}^2$$  \hspace{1cm} \text{(8.68)}

A plot of $W_{sp}$ as a function of the energy separation $E_{f_i}$ is shown in Fig.(8.19). The oscillator strength was set to $f_{12} \approx 0.96/m^* \text{ between two states in a well.}$ We have taken the effective mass of the Ga$_{0.47}$In$_{0.53}$As : $m^* = m_w \approx 0.042$. The spontaneous emission rate is proportional to $E_{f_i}^2$.

The quantum efficiency of the spontaneous emission is given by: $\eta = W_{sp}/(W_{sp} + W_{nr})$, where $W_{nr}$ is the total scattering rate of non-radiative processes. Since the latter are of the order of the picosecond in mid-infrared quantum cascade lasers, the radiative efficiency of the spontaneous emission is very low: $\eta \approx 3.3 \cdot 10^{-5}$ for $E_{f_i} = 167 \text{ meV}$. 


Chapter 8. Light interaction modeling in quantum cascade lasers

Figure 8.19: Spontaneous emission rate as a function of the separation energy between the lowest states of quantum well of Ga$_{0.47}$In$_{0.53}$As.

However the measurement of spectrally resolved luminescence (spontaneous emission) enables the measurement of the linewidth of the transition. The latter is not affected by spontaneous emission as shown by the value of the quantum efficiency. These measurements were used in order to estimate the dephasing time $\tau_{i,e}$ at the injection and extraction resonances for the far-infrared structure discussed in (6.5) and completed below.

We first consider simple rate equations and the simplest model of the laser field in a cavity. If we consider the upper laser level (3) and the lower laser level (2), we can write the rate equations:

$$ \dot{n}_3 = \frac{J}{q_0} - \frac{n_3}{\tau_3} - S g_c (n_3 - n_2) $$
$$ \dot{n}_2 = \frac{n_3}{\tau_{32}} - \frac{n_2}{\tau_2} + S g_c (n_3 - n_2) $$

$$ \frac{dS}{dt} = \frac{c}{n_r} \left[ (g_c (n_3 - n_2) - \alpha_{tot}) S + \beta \frac{n_3}{\tau_{sp}} \right]. $$

The population of the upper laser state is given by $n_3(t)$, while the lower state population is given by $n_2(t)$. The upper laser state is pumped by a current density $J$. Its total lifetime is $\tau_3$ while the scattering time corresponding to the transition $3 \rightarrow 2$ is given by $\tau_{32}$. The lower state has a total lifetime of $\tau_2$ (we assume that $\tau_{23} = \infty$).
The laser field is modeled by a *photon flux density* $S$ at a defined energy $\hbar \omega_0$ fixed by the cavity mode. It is the photon flux per period and per unit of active region width. The latter refers to the width of a ridge waveguide for example. The stimulated emission and the absorption are modeled by the scattering terms $-S g_c (n_3 - n_2)$ and $+S g_c (n_3 - n_2)$, where $g_c$ is the gain cross-section. It is worth to remark that a first-order gain model was considered for this example as the gain is proportional to the net population inversion.

The gain cross-section accounts for the modal overlap $\Gamma$ factor between the optical mode, in the cavity of refractive index $n_r$, and the active region. The modal overlap factor satisfy $0 \leq \Gamma \leq 1$ and is defined by:

$$\Gamma \equiv \frac{\int_{\text{active region}} d^3r E^2(r)}{\int d^3r E^2(r)}, \quad (8.70)$$

where $E(r)$ is the electric component of the cavity mode.

The bare gain cross-section (normalized to the period length $L_p$) as derived from Eq.(8.56) needs therefore to be multiplied by $\Gamma$. The gain multiplied by the overlap factor is the modal gain, as $\Gamma$ accounts for the fraction of the light (in the optical mode) that is amplified by the gain medium.

The photon-flux density $S$ is the solution of a time dependent equation in Eq.(8.69) that includes the spontaneous emission. The latter is multiplied by a factor $\beta$ which is the ratio of the intensity of spontaneous emission injected in the laser mode. The origin of the laser oscillation is clearly the spontaneous emission. However, the laser field is driven by the term: $(g_c (n_3 - n_2) - \alpha_{\text{tot}}) S$ that actually causes light amplification or attenuation. The fraction of the gain that contributes to the amplification is given by: $g_c (n_3 - n_2) - \alpha_{\text{tot}}$, where $\alpha_{\text{tot}}$ are the total losses. In steady-state operation, the modal gain is therefore clamped to the total losses. If the latter overcome the gain, the photon flux is attenuated.

The total losses consist of two terms: the cavity losses $\alpha_{\text{w}}$ and the losses due to inter-subband absorption in the active region $\alpha_{\text{isb}}$.

We have to be very careful. The cavity losses $\alpha_{\text{w}}$ are somewhat well-defined. They are the losses due to the hollow cavity, where the active region has been replaced by a dielectric medium. They include the mirrors (or reflecting boundaries) losses and the confinement losses.

In usual ridge waveguide configurations it is however useful to distinguish the hollow waveguide losses $\alpha_{\text{wg}}$ that essentially consist in the losses due to the confinement layers (cladding layers), from the mirror losses $\alpha_{\text{m}}$. This distinction is useful as the waveguide losses are usually held constant (given by semiconductor processing), while the mirror
losses can we varied by choosing different coatings. For a ridge waveguide configuration we have: \( \alpha_w = \alpha_{wg} + \alpha_{m,bc} + \alpha_{m,fc} \), where we have distinguished the losses due to the front-facet (fc) from the losses due to the back-facet (bc).

If we now consider inter-subband losses \( \alpha_{isb} \), the latter have a meaning if the gain cross-section \( g_c \) only includes the laser states. In this case, we have to add the losses due to inter-subband absorption (in the injector typically) at the frequency of the cavity. However if the gain cross-section includes all possible transitions in the heterostructure, the (additional) inter-subband losses are zero: \( \alpha_{isb} = 0 \). The usual free-carrier absorption in bulk materials is translated for heterostructures into inter-subband absorption. No supplementary terms are necessary as shown in \[112\].

In simpler models, the inter-subband absorption is replaced by the absorption caused by dopants in the active region. The corresponding losses are computed by a Drude model. However if inter-subband absorption (with second-order terms) was computed between all states, this additional contribution is redundant and gives over-estimated losses. Moreover, it is worth to underline that the Drude model is a very bad approximation for computing the losses in the active region as the band-structure is not taken into account and therefore parasitic transitions (in the injector region) tuned at the laser transition-energy cannot be modeled.

Here we do not spent time in computing the steady-state solution of Eq.(8.69) and we jump directly to the modeling of \( S \) for far-infrared single-quantum-well structures. The computation of the optical power is postponed to (8.4).

### 8.3.1 The photon-flux density

We consider the four-state (0-dimensional) model developed in (6.5) in the density-matrix formalism. This system is illustrated in Fig.(6.11). We add a photon flux density \( S \) that causes stimulated emission and absorption between the upper (2) and lower (1) laser states. The equations of motion completed with the flux density \( S \) are given by Eq.(6.83) under the transformations: \( \dot{\rho}_{22} \rightarrow \rho_{22} + S g_c (\rho_{22} - \rho_{11}) \) and \( \dot{\rho}_{11} \rightarrow \rho_{11} - S g_c (\rho_{22} - \rho_{11}) \). In this simple model we consider only the first-order gain contribution. We have:

\[
\begin{align*}
\dot{\rho}_{gg} &= -i \Omega_i (\rho_{2g} - \rho_{g2}) + \frac{\rho_{uu}}{\tau_u} \\
\dot{\rho}_{22} &= i \Omega_i (\rho_{2g} - \rho_{g2}) - \frac{\rho_{22}}{\tau_2} - S g_c (\rho_{22} - \rho_{11}) \\
\dot{\rho}_{11} &= -i \Omega_e (\rho_{u1} - \rho_{1u}) + \frac{\rho_{22}}{\tau_2} + S g_c (\rho_{22} - \rho_{11}) \\
\dot{\rho}_{uu} &= i \Omega_e (\rho_{u1} - \rho_{1u}) - \frac{\rho_{uu}}{\tau_u}
\end{align*}
\]

(8.71)
8.3. Computation of light-current curves in the far-infrared

\[
\dot{\rho}_{g2} = -i\Omega_i(\rho_{22} - \rho_{gg}) - i\Delta_i \rho_{g2} - \frac{\rho_{g2}}{\tau_{\perp,i}}
\]
\[
\dot{\rho}_{2g} = i\Omega_i(\rho_{22} - \rho_{gg}) + i\Delta_i \rho_{2g} - \frac{\rho_{2g}}{\tau_{\perp,i}}
\]
\[
\dot{\rho}_{1u} = -i\Omega_e(\rho_{uu} - \rho_{11}) - i\Delta_e \rho_{1u} - \frac{\rho_{1u}}{\tau_{\perp,e}}
\]
\[
\dot{\rho}_{u1} = i\Omega_e(\rho_{uu} - \rho_{11}) + i\Delta_e \rho_{u1} - \frac{\rho_{u1}}{\tau_{\perp,e}}
\]

with \( \rho_{gg} + \rho_{22} + \rho_{11} + \rho_{uu} = 1 \).

We still have: \( \dot{\rho}_{gg} + \dot{\rho}_{22} + \dot{\rho}_{11} + \dot{\rho}_{uu} = 0 \). The photon-flux density is here fixed and arbitrary. The threshold condition is implemented below. The gain-cross section is evaluated here at the maximum of the gain and it includes the modal overlap \( \Gamma \).

The upper-state lifetime is reduced by \( S \):

\[
\tau_2(S) = \frac{\tau_2}{1 + g_c S \tau_2}.
\] (8.73)

However we cannot simply use the solution for \( S = 0 \) and substitute \( \tau_2(S) \) into it, as absorption gives a finite lifetime between the lower and the upper laser levels. Nevertheless the calculations are quite simple and we give the results only. We have to substitute populations in the conservation of the norm: \( \rho_{gg} + \rho_{22} + \rho_{11} + \rho_{uu} = 1 \). At the end, we can express the current density with an effective transport time \( \tau_{\text{eff}}(S) \) that depends on \( S \) [83]. We have:

\[
J(S) = \frac{g_0 N_s \tau_{\text{eff}}^{-1}(S)}{\tau_{\text{eff}}(S)}
\]

\[
\tau_{\text{eff}}(S) = 2 \left( \frac{\tau_u + \tau_2 + g_c S (\tau_u + T_e)}{1 + g_c S \tau_2} \right) + T_e + T_i.
\] (8.74)

\[
T_{i,e} = \frac{1 + \Delta_{i,e}^2 \tau_{\perp,i,e}^2}{2 \Omega_{i,e}^2 \tau_{\perp,i,e}^2}
\]

Where \( T_{i,e} \) is the tunneling time at injection and extraction resonances. The photon flux-density clearly affects the effective transport time.

We can consider various coupling regimes between radiative and non-radiative transports. In particular a strong coupling regime where \( g_c S \tau_2 \gg 1 \). In this limit the lifetime of the upper state is given by: \( \tau_2(S) \approx (g_c S)^{-1} \), which means that the non-radiative lifetime of the upper-state can be neglected against the stimulated emission lifetime: the system is driven by radiative scattering and no more by non-radiative lifetimes. We have: \( \tau_{\text{eff}} \approx 2(2\tau_u + T_e) + T_e + T_i \). In this special system where the extraction mechanism is based on resonant tunneling, the photon flux density can reduce the upper-state lifetime but not
Chapter 8. Light interaction modeling in quantum cascade lasers

the extraction lifetime. The above limit expression does not depend on \( S \) and \( \tau_2 \) and the amount of current is limited by the extraction time. We will however not observe this regime in the investigated far-infrared sample N471.

It is worth to remark that in a system that relies on incoherent scattering for the depletion of the lower state, the strong coupling regime with the light implies that the maximal current density is driven by the radiative lifetime \( \tau_2 \approx (g_c S)^{-1} \). In this case we therefore expect that the current density can be largely increased with respect to its zero field \((S = 0)\) value. Before we focus only on the resonant injection/extraction model, we should consider the case of incoherent extraction in a strong coupling regime: \( g_c S \tau_2 \gg 1 \). We consider Eq.(6.72) for the two-state model. We have: \( \tau_1 \approx \tau_2 \approx g_c S \) and therefore \( J = 0 \). This result is the consequence of the reduced phase space in this system. If non-radiative lifetimes are negligible, the current is therefore zero by symmetry between the stimulated emission and the absorption.

If we now consider a system where the current is given by the total lifetime of the upper laser-level (this is formally the limit \( \tau_1 \rightarrow \infty \) for the two-state model), then, the current is driven by the lifetime \((g_c S)^{-1}\). If this system was already in a strong coupling regime without the optical field (6.4.6), the maximal current is given by: \( J(S) \approx q_0 N_s \frac{1}{2} g_c S \). The current is only driven by the photon flux density \( S \). This is however a very rough model.

Now we focus on the resonant system in Eq.(8.74). The current density is given as a function of \( S \). However \( S \) is arbitrary as we do not have imposed a threshold condition for it. We need to solve \( S \) self-consistently. If we look back to Eq.(8.69). The equation for \( S \) can be expressed as: \( \dot{S} = \gamma S + C_0 \). The coefficient \( \gamma \) is given by: \( \gamma = \frac{\varepsilon}{n_r} (g_c (n_3 - n_2) - \alpha_{\text{tot}}) \).

It contains the threshold condition and depends on time. The term \( C_0 \) contains the contribution from spontaneous emission. We assume the latter is constant. We first assume that \( \gamma \) is also constant with respect to time. This is not a good approximation, but we do not want to carry integrals in the development. At least this approximation holds when \( g_c S \) is low enough for a time interval of several \( C_0^{-1} \). In this case, the general solution is given by:

\[
S(t) = S(0)e^{\gamma t} + \frac{1}{\gamma} \left( 1 - e^{-\gamma t} \right) C_0
\]

We assume initially that \( S(0) = 0 \). The time evolution is therefore given by the inhomogenous solution. If the medium is amplifying \((\gamma > 0)\), we can expand the exponential in the short time limit \( \tau \), and we have: \( S(\tau) \approx \tau C_0 \). As expected the photon-flux density increases with time and the slope is given by the spontaneous emission rate in the optical mode of the cavity.

However we do not want to treat the time-dependent problem. We remark that the increase
of \( S \) will decrease the population inversion and therefore the amplification coefficient of the system. We therefore expect that the system will be stable when the modal gain is equal to the losses: \( \gamma \equiv 0 \). In this case, by neglecting the spontaneous emission term: \( \dot{S} = 0 \).

If we compute the population inversion as a function of \( S \), the steady-state solution is obtained when \( S \) satisfy:

\[
g_c (n_3(S) - n_2(S)) \equiv \alpha_{\text{tot}}
\]

(8.76)

We can compute the population inversion in the four-state resonant model as a function of \( S \). From the equation of motion for \( \rho_{22} \), we have in steady-state:

\[
\rho_{22} = \frac{\tau_2}{1 + g_c \tau_2 q_0 N_s} + \frac{g_c S \tau_2}{1 + g_c \tau_2} \rho_{11}.
\]

(8.77)

And \( \rho_{11} = \rho_{uu} + T_c \frac{J}{q_0 N_s} \) together with \( \rho_{uu} = \tau_u \frac{J}{q_0 N_s} \), and we obtain the population inversion \( \Delta N(S) \equiv \rho_{22} - \rho_{11} \):

\[
\Delta N(s) = \frac{1}{1 + g_c \tau_2} \frac{\tau_2 - (\tau_u + T_c)}{\tau_{\text{eff}}(S)}
\]

(8.78)

where \( \tau_{\text{eff}}(S) \) is given in Eq.(8.74). We have to solve the equation for \( S \):

\[
g_c \Delta N(S) = \alpha_{\text{tot}}.
\]

(8.79)

For this simple system this equation has an analytic solution:

\[
g_c S = \frac{1}{\alpha_{\text{tot}}} \frac{g_c (\tau_2 - (\tau_u + T_c)) - \alpha_{\text{tot}}(4 + T_i + T_e)}{2(1 + \tau_2)(\tau_u + T_c) + \tau_2 T_i}.
\]

(8.80)

The emission region is defined by \( S \geq 0 \). The boundary points of a region where \( S \geq 0 \) correspond to threshold currents. The lowest current is the usual threshold current density and the next one is called extinction threshold current.

It this simple system an analytic solution exists. However in a general configuration, we have to search numerically (typically with a bisection algorithm) the value of \( S \) that satisfy: \( g_c \Delta N(S) = \alpha_{\text{tot}} \). In (8.4) we use this technique for the computation of \( S \).

### 8.3.2 Impact of resonant tunneling on emitted power

In the previous section, we have seen that regions where \( \Delta N \geq \alpha_{\text{tot}}/g_c \) allow solutions with \( S \geq 0 \). We have computed for sample N471 described in (6.5) the value of \( S \) in the region where the modal gain exceed the losses. The situation is illustrated in Fig.(8.20).
Chapter 8. Light interaction modeling in quantum cascade lasers

Figure 8.20: The gain is clamped to the losses by the condition: $g_c \Delta n = \alpha_{\text{tot}}$. The roll-over in the population is therefore converted into a roll-over on the photon-flux density $S$. The system shows two threshold currents and the extinction threshold current is lesser than the maximal current.

The population roll-over is translated into a roll-over on the light characteristic. The values of both threshold currents and of the maximal current depend strongly on the lifetimes $\tau_2$ and $\tau_u$, but also on the dephasing times $\tau_{\perp,i(e)}$. A typical measurement is shown in Fig.(8.21).

The dephasing times are first roughly deduced from a luminescence measurement shown in Fig.(8.22). The linewidth corresponds to a typical dephasing time of $\tau_{\perp} \approx 1.3$ ps. This gives a rough estimate of the dephasing time as the non-radiative lifetimes are one order of magnitude higher and therefore do not contribute significatively to the linewidth.

The lifetimes times are first estimated from the maximal current $J_{\text{max}}$ and the threshold current density $J_{\text{th}}$. In a crude approximation $J_{\text{max}} \approx g_0 n_s/2(\tau_2 + \tau_u)$ which yields an upper bound for the sum: $\tau_2 + \tau_u < 26$ ps. For an extraction mechanism based on incoherent
Figure 8.21: LJV curves for N471 sample taken at 8 K in continuous wave. The sample was processed in a double-metal ridge-waveguide cavity of 150 μm width and 0.94 mm length. The predicted effect due to the misalignment of resonant extraction is clearly observed. The laser action starts at $J_{\text{th}} \approx 29 \, \text{A/cm}^2$, stops at $J_{\text{ext}} \approx 92 \, \text{A/cm}^2$ and the maximal current is reached at $J_{\text{max}} \approx 120 \, \text{A/cm}^2$. The heating of the sample cannot be incriminated as this feature is also observed in pulsed mode for low duty cycles.

scattering we have: $J_{\text{th}} \approx (q_0 \alpha_{\text{tot}}/g_c)/(\tau_2 - \tau_u)$, which yields an approximative value: $\tau_2 - \tau_u \approx 3 \, \text{ps}$. In these calculations, the total losses were measured: $\alpha_{\text{tot}} = 12 \, \text{cm}^{-1}$, for a double-coated, double-metal ridge waveguide of 1 mm.

The experimental technique is based on measurements of the threshold current as a function of the ridge waveguide length [38]. The deduced losses that include intrinsically the inter-subband losses are relevant here, as we model the gain for laser states only.

The gain cross section was computed for the laser doublet: $g_c = 2.37 \cdot 10^{-8} \, \text{cm}^{-1}/\text{cm}^2$, assuming a linewidth of $\gamma = 1.0 \, \text{meV}$ measured from spontaneous emission. The linewidth also gives an estimation for the dephasing times ($\tau_|| \approx 1.3 \, \text{ps}$) required by the resonant
Figure 8.22: Luminescence measurements taken at 4.2 K. With increasing duty cycle a narrowing is observed before lasing. The transition energy is \( \approx \) 15.5 meV and the broadening of the luminescence (without narrowing) gives a linewidth of \( \approx \) 1.0 meV. The linewidth is narrow as expected from a vertical transition in a single-quantum-well structure.

Based on these gross values, the relaxation times were then optimised to fit the experimental data. Refined values were obtained: \( \tau_2 = 12.2 \) ps, \( \tau_u = 8.9 \) ps for population relaxation times and \( \tau_{|i|} = 1.6 \) ps, \( \tau_{|e|} = 1.2 \) ps for dephasing times. As expected from the low current density in these samples the relaxation times are about a ten of picoseconds and the spontaneous emission linewidth is therefore dominated by the dephasing time.

The model fully explains the laser characteristics, but an even more direct experimental evidence is obtained by comparing samples with different total losses. Since the population inversion and thus the gain between laser states is driven by resonant tunneling at the
8.3. Computation of light-current curves in the far-infrared

Figure 8.23: Differential resistance of samples: with single facet coated (dotted) and double-coating (solid). Upper state lifetime (inset) reduced by stimulated emission. Arrows underline the derivative at threshold current (33 A/cm$^2$) and when the laser stops (112 A/cm$^2$).

extraction barrier, the current range where light is on will be increased by lowering the total losses. As the threshold condition is met for less injected electrons, the threshold current density is reduced, while at the same time, the laser action can be sustained up to higher currents. As shown in Fig. (8.23), the differential resistance was measured for samples with different coatings. A reference sample was first measured in a ridge configuration, with a single coated facet, then the other facet was coated in order to reduce the losses. As expected the light dynamic range is clearly improved: the threshold current density is reduced from 38 A/cm$^2$ to 33 A/cm$^2$, while the laser stop at a higher current density: 107 A/cm$^2$ for the single coated facet and 112 A/cm$^2$ with the double coating.

The variation of the differential resistance is negative both at laser threshold and when laser stops. This is explained by looking at the sign of the derivative of the upper state lifetime, as shown in the inset of Fig. (8.23). At threshold, the upper state lifetime goes down abruptly from its zero-field value, decreasing afterwards from 12.2 ps to a minimum of 10.4 ps. As the optical field intensity falls when the extraction becomes less efficient, the lifetime re-increases to 12.2 ps, where it is again abruptly reduced to the zero-field value when laser stops.
8.4 Coupled optical field intensity and effective transport model

We implement the computation of the photon flux density $S$ at a fixed photon energy $\hbar \omega_0$ in the effective model we have developed in (7). The predictions of the completed model were published in [94].

8.4.1 Method for computing the photon-flux density $S$

We consider the modal gain $g_{ij}(\hbar \omega)$ between a pair of subbands $(i, j)$ inside the active period. In the transport model developed in this work we have to assume that the pair of subbands is inside a period (or a sub-period if more than one coupling barrier is considered). Technically we are able to compute the diagonal gain between periods. The latter is however not well-defined in a tight-binding basis and it is numerically very weak as the overlap between the wavefunctions are in this representation very low. For modeling the tunneling-assisted gain across a coupling barrier we have to implement the model developed by E. Dupont in [25], for example. At the time of writing this model was however not been implemented. The effect of $S$ on the net populations $n_i$ and $n_j$ can be modeled by:

\[
\dot{n}_i = \ldots - g_{ij}(\hbar \omega_0)S + g_{ji}(\hbar \omega_0)S \quad (8.81)
\]

\[
\dot{n}_j = \ldots + g_{ij}(\hbar \omega_0)S - g_{ji}(\hbar \omega_0)S \quad (8.82)
\]

If we assume that $g_{ij}$ is a linear form in the intra-subband carrier-distributions $f^i(k)$ in the subbands, we can write:

\[
g_{ij}[f^i; f^j] = g_{ij}[f^i; 0] + g_{ij}[0; f^j]. \quad (8.83)
\]

This property is valid for the first-order gain model and for the second-order gain model. For the first-order model this is even true for the net populations as shown in Eq.(8.23). For the second-order gain model, we have in Eq.(8.56):

\[
g_{ij} \equiv g(\omega_0) = \frac{g_0^2 |z|^2 (\hbar \Delta)^2}{\varepsilon_0 n_r c h^2 \omega} \sum_k \frac{\gamma(k)(f_{ii}(k) - f_{jj}(q_+)) + \gamma(k)(f_{ii}(q_-) - f_{jj}(k))}{(\hbar \Delta - \hbar \omega)^2 + (2\gamma(k))^2}. \quad (8.84)
\]

Even if the populations are evaluated at special wave-vectors, the expression is linear and homogeneous in $f_{ii}$ and $f_{jj}$ as there is no product of the distributions: $(f_{ii})^2$, etc. The Eq.(8.83) therefore holds in both cases. This property enables us to define gain cross-sections as:

\[
g_{ij}^{c,i} = \frac{g_{ij}[f^i; 0]}{n_i} \quad \text{and} \quad g_{ij}^{c,j} = \frac{g_{ij}[0; f^j]}{n_j} \quad (8.85)
\]
8.4. Coupled optical field intensity and effective transport model

when \( n_i \neq 0 \) and \( n_j \neq 0 \). Else we set \( g_{ij}^{c,i} = 0, \ g_{ij}^{c,j} = 0 \). The gain between upper subband \( i \) and lower subband \( j \) can therefore be rewritten as:

\[
g_{ij} = g_{ij}^{c,i} n_i + g_{ij}^{c,j} n_jan{8.66}
\]

This linear structure enables us to write the effect of the photon-flux density as an additional term in standard rate equations:

\[
\begin{pmatrix}
\dot{n}_i \\
\dot{n}_j
\end{pmatrix} = \left( \ldots \right) + S \begin{pmatrix}
g_{ij}^{c,i} + g_{ji}^{c,i} & g_{ij}^{c,i} - g_{ji}^{c,i} \\
g_{ij}^{c,j} + g_{ji}^{c,j} & g_{ij}^{c,j} - g_{ji}^{c,j}
\end{pmatrix} \cdot \begin{pmatrix}
n_i \\
n_j
\end{pmatrix}.an{8.67}
\]

However this linearization gives accurate results only if it is embedded into a self-consistent loop. We first compute the populations with the transport model at zero field \( (S = 0) \). We then compute the effective gain cross-sections with the above linear model and an initial value for the photon-flux density \( S \). The system is then solved with the additional term given in Eq.\( (8.67) \): this requires sub-selfconsistent iterations that compute stable populations and the electronic temperature as required by the effective transport model developed in \( (7) \).

It is important to mention here that we have linearized the system only locally with Eq.\( (8.67) \). As the solution is self-consistent all relevant transport and optical quantities are computed iteratively, like, for example, the optical linewidths. We are therefore able to predict non-linear effects, like \textit{gain compression}, where the decreased lifetime of the upper laser-state due to stimulated emission causes a broadening of the gain curve as the linewidth is defined self-consistently \( (5.2.3) \).

Finally we obtain new populations that depends on the value of \( S \). Here we evaluate the total spectral modal gain \( G(\hbar \omega) \) at the energy \( \hbar \omega_0 \). The latter is the sum of each possible transitions:

\[
G(\hbar \omega) = \Gamma \sum_{(i,j)} g_{ij}[f_i, f_j](\hbar \omega)an{8.68}
\]

We evaluate the threshold conditions given by the total losses \( \alpha_{\text{tot}} \) of the hollow cavity as discussed in \( (8.3) \):

\[
\delta g_{th} = G(\hbar \omega_0) - \alpha_{\text{tot}}an{8.69}
\]

If the value of \( \delta g_{th} \) is greater than zero \( (\delta g_{th} > 0) \) it means that the photon flux \( S \) is too low, while for \( \delta g_{th} < 0 \) it is too high. The equality representing the actual photon-flux density in the cavity. For searching the zero of \( \delta g_{th} \) we have implemented an efficient algorithm that first bisects the region where \( \delta g_{th} \) is expected to be zero and then use a Brent method \( [77] \) to converge on the solution. In the above method we implicitly assumed that for a given electric field, the function \( \delta g_{th}(S) \) has a single root.
Once the photon-flux density is computed at a given energy $\hbar \omega_0$, we can extract the power measured from one facet of a laser ridge waveguide. If $N_p$ is the number of periods in the structure and $w_r$ is the width of the laser ridge, the optical power $P_{opt}$ extracted from one facet of the ridge is given by:

$$P_{opt} = N_p \hbar \omega_0 S w_r \alpha_{fm}$$

Where $\alpha_{fm}$ is the mirror losses given by a facet of reflectivity $R$, assuming we are extracting the light from one facet. For both facets, we have: $\alpha_m = -\log (R_1 R_2) / 2L$ where $L$ is the length of the waveguide and $R_{1,2}$ the reflectivity of the facets. Here, for standard computations, we consider $R_1 = 1$ (high-reflection coated back-facet) and $R = 0.25$ for the front facet. Typically for a ridge of 3 mm, we have: $\alpha_m \approx 2.31 \text{ cm}^{-1}$.

We can also normalize the internal losses $\alpha_{tot}$ to a typical value for mid-infrared waveguides: $L = 3 \text{ mm}$, with high-reflectivity back-facet ($R=1.0$) and $R=0.25$ for the front facet. If we include the modal overlap factor $\Gamma \approx 0.68$ in the total losses, we get the typical value of $\alpha'_{tot} \approx 6 \text{ cm}^{-1}$. The threshold condition therefore reads: $G(\hbar \omega_0) - \alpha'_{tot} = 0$, as we have incorporated the modal overlap in the definition of the total losses.

### 8.4.2 Computation of the spectral gain

The total spectral gain is computed by summing all possible transitions between the subbands inside a period (or a sub-period), possibly taking into account the diagonal gain between periods. The Eq. (8.88) summarize this computation by taking into account the modal overlap factor $\Gamma$. This computation therefore incorporates the resonant losses due to the tail of absorption in the injector region. Regarding the losses that have to be taken into account in the calculation of the photon-flux density, we have set: $\alpha_{isb} \equiv 0$.

For each pair of subbands we have assumed thermal intra-subband distributions with a uniform electronic temperature. The gain can be computed by using the first- or the second-order gain formula. The linewidth is computed by evaluating Eq. (5.2.3) by taking into account all scattering mechanisms described in (5). In the evaluation of the linewidth we may integrate over the populations of each subband as $\Gamma_{\nu \mu}(k)$ depends on $k$. However to reduce the computational effort we have simply evaluated the optical line at the thermal energy $k_b T_e$ corresponding to the electronic temperature $T_e$. This may require a better implementation in the future.

It is worth here to point out a limitation of the gain model used here and discussed briefly in the previous section. We are considering potential regions coupled by barriers. The gain is therefore evaluated inside the regions and possibly between them by direct radiative
scattering. However if we consider a left region that injects carriers resonantly into a right region where an optical transition occurs, we expect for thick enough barriers that the gain consists in a single resonance found at the transition-energy between states in the right region. Nevertheless if the barrier is made thinner we expect that at some point the coupling energy will be largely greater than the depashing energy and that the gain resonance will be split into two resonances where the transition energies are given between the upper/lower state of the coherent doublet and the lower state in the right region. Our gain model does not account for this situation accurately and therefore overestimates the gain for thin barriers. However there exist an optimal width of the barrier where the gain is maximal as shown by [25, 57]. The gain models presented in these papers can formally be implemented in our transport model, but this has not been done so far.

8.4.3 Light-current-voltage curves of a two-phonon structure

We apply the model to a reference two-phonon quantum cascade laser [116]. The lattice heating can be neglected as all measurements were taken in pulsed mode. Regarding the simulation we have taken typical parameters for the various scattering mechanisms. The interface roughness is modeled [103] with a value of $\Lambda = 90$ Å for the in-plane correlation length of the steps, $\Delta = 1.2$ Å for the step height, and $\kappa = 15$ Å for the correlation length between interfaces. These values are not empirical but were deduced from measurements in [101, 116] and discussed in (5.3.1). They influence weakly the current-voltage characteristic but can be crucial in the threshold current determination as interface roughness is the main broadening mechanism in mid-infrared quantum cascade lasers. The LO-phonon energy is $\hbar \omega_{LO} = 32$ meV. The computation time for the current and the photon flux density is about 150 seconds per point (at a given electric field) on current hardware. We have parallelized the computation of the light-current-voltage curves and we actually run it on a cluster.

In order to validate our model we have simulated the temperature dependence of the threshold current. The threshold current follows usually $I_{th} = I_0 \exp(T/T_0)$. As a first step, we have simulated the current-voltage curve using the global kinetic balance model. The result is shown in Fig.(8.24). As in [96], an excellent agreement exists between experiment and theory. But the simulated threshold current is slightly higher than the experimental value. The optical power is also lower. The discrepancy increases with temperature.

The global kinetic balance model is too pessimistic because it overestimates the electronic temperature of most subbands: at the injection resonance ($J = 5.2$ kA/cm$^2$) with a

\footnote{Intel Core Duo at 2 Ghz}
lattice temperature of 300 K, the electronic temperature is found near 600 K. This can be explained by looking closer to the laser transition. An electron injected in the upper laser state with no excess of kinetic energy will perform a non-radiative transition to the lower laser state by emitting an optical phonon (inelastic) or by elastic scattering. In any case it will end in the lower subband with a high in-plane momentum. The model will then try to represent these hot electrons with a thermal distribution, forcing a strong heating of all subbands. The overestimation of the electronic temperature of the injector subbands causes backfilling in the active wells while it enables an optical phonon absorption from the upper laser state to excited states. Both processes reduce the population inversion and therefore explain the increase of the threshold current.

In Fig. (8.25) we show the electron distribution versus the energy at the injection resonance for two opposite cases: when the global kinetic balance is solved, and when the electronic temperature is forced to the lattice temperature.

The latter case considers that the electron-lattice interaction is strong enough to thermalize the electrons (they essentially reside in the injector region) to the lattice temperature. This reproduces plausibly the situation of the structure showed here. In fact, we can
8.4. Coupled optical field intensity and effective transport model

![Figure 8.25: Subband populations versus energy. The dashed line represents the case where the electronic temperature is forced to the lattice temperature. The solid line is the distribution found by solving the kinetic energy balance model.](image)

estimate the average time $\bar{\tau}$ spent by an electron in the active wells region. For an applied electric field of 34 kV/cm in the middle of the dynamic range, we compute the sum of the carrier density in the subbands of the active wells region: $n_a \approx 0.26 \cdot 10^{11}$ cm$^{-2}$, and the corresponding current density: $J \approx 2.6$ kA/cm$^2$. Using $J = q_0 n_a / 2\bar{\tau}$ in [88], we have $\bar{\tau} \approx 1.6$ ps. During this time the electron can emit approximately 8 optical phonons [28, 31], using 200 fs as the emission rate of bulk optical phonons. It allows the electron to lose $\approx 256$ meV of kinetic energy, covering largely the energy gap formed by the optical transition and the phonon resonances designed between the lower states of the active wells. Therefore, most electrons have no excess kinetic energy when they are extracted to the injector region. In the injector itself, the transport is achieved by direct scattering between a dense collection of states that allows an efficient thermalization. Therefore we kept the subbands to the lattice temperature as shown in Fig. (8.24) at 300 K. A very good agreement is found. It is important to notice that no fit parameters were used.

We now test the model on a larger range of temperatures by computing the $T_0$ parameter.
Figure 8.26: (a) Threshold current versus lattice temperature. Simulated values are represented by circles for the electron-lattice thermalization, by triangles for the kinetic energy balance model and by squares for the experiment. The solid lines are fits with the $T_0$ function. (b) Comparison between theory and experiment of light-current curves for the two extreme temperatures.
8.5 Towards the automatic optimization of quantum cascade lasers

The latter is robust and will show if we have chosen the right thermal model. In Fig. (8.26a), the threshold current was plotted versus the lattice temperature and then fitted with the $I_0 e^{T/T_0}$ function. The value extracted from measurements is $T_0 = 174$ K. The simulated value with the kinetic energy balance model is $T_0 = 111$ K while the result for equal lattice and electrons temperatures is $T_0 = 155$ K, in a better agreement with the experiment.

We expect a different situation in strained compensated structures where the electronic temperature was measured [106] clearly above the lattice temperature. The simplification made here does not hold anymore as the energy-gap between the upper laser state and the injector ground state is much larger.

In Fig. (8.26b) we show the light-current curves for the lowest and the highest lattice temperatures. The agreement is excellent in terms of threshold currents and slope efficiencies.

8.5 Towards the automatic optimization of quantum cascade lasers

The automatic optimization of quantum cascade structures has been discussed conceptually [70]. The basic concepts are clear: we need to associate with each structure, represented by the layer sequence or some effective parameters, a merit function. An optimum is represented by an extremum of the merit function. However if we consider the problem in its generality we have a scalar function that depends on $M$ layer thicknesses, an electric field $F$ and a temperature $T$. The latter gives the LO-phonon bath temperature. For a typical mid-infrared laser $M \sim 25$ and therefore we have a function of roughly 27 variables. As the function is known only numerically it is essentially impossible to perform usual differential analysis. It is however true that we can apply relatively strong constraints on the parameters. For example, the domain of variation of each layer thickness can be limited to 0.2 nm to 50 nm. Anyway it is extremely difficult (even impossible) to show that the merit function has a global extremum and to show that this extremum is stable and not a kind of cusp or simply numerical noise.

There exists fortunately quite general algorithms, like thermal annealing [77], that enable the minimization of a merit function in spaces with a high dimension. The next problem is the time required for running such optimization method on actual quantum cascade structures. If we can only compute a few points (required for building the standard simplex) in parallel, it is worth to figure that whether we optimize one (or a few) layer in the structure or the optimization will never converge in a decent time.

We may therefore be interested in some kind of genetic algorithms. If we consider an initial
structure and we perform a sequence of random variations of the sicknesses that mimic the spontaneous mutations in nature, we evaluate the merit function on this population and and we select a few better individuals that we will retain for building the next generation. The merit function there implements the Darwin’s selection rule. We need to take care to select a large enough pool of individuals for building the next generation to basically avoid the degeneration which here represents the convergence on some unstable extremum of pathological structure that shows in fact the limits of the underlying physical model.

In nature we also need a cross-over function that models the reproduction of the individuals. This generation mechanism is competing the random variation. The first tends to fix a character while the other try to explore the available phase space. These two mechanisms enable the creation of a new generation. We expect that after several generations we will obtain a population spread in the vicinity of a stable maximum of the merit function.

From this ideal scenario we have retained a few concepts and we have implemented them in order to try to optimize mid-infrared quantum cascade lasers. The first very important choice is the merit function. Here we have chosen the wallplug efficiency \( \eta_{wp} \). It is the ratio between the optical power and the electrical power. We define it as:

\[
\eta_{wp} = \frac{S \hbar \omega_0 \alpha_{wp}}{F_0 L_p J},
\]

where \( S \) is the photon-flux density with photons of energy \( \hbar \omega_0 \), \( F_0 \) is the applied electric field, \( L_p \) the period length and \( J \) the current density. The losses \( \alpha_{wp} \) represent usually the mirror losses that give the extracted power from one facet. However we want to compute here the total power from the ridge and \( \alpha_{wp} \) therefore represent the losses from both facets. The definition of this number nevertheless depends on the definition we have chosen for the total internal losses. Anyway this number is somewhat a fixed parameter. We can fix it and compute the optical power always with the same convention and \( \eta_{wp} \) will be anyway a well-defined merit function.

On the physical side is we let the algorithm optimize this function we will end with power-efficient structures which is the first interesting optimization [27].

We do not have completely implemented the genetic side of the algorithm as we are missing the cross-over operation. In some sense we are doing random optimization of structures. What is retained from the genetic algorithm is the initial population chosen as a random variation of an initial structure and the selection rule with the merit function as we retain a pool of structures from generations to generation and not only the best one.

To give more details on the structure of the algorithm, each generation has typically 32768 individuals. The genome consist of the layer sequence and the electric field. However we constraint the latter to a finite interval. The variation range of the electric field
is typically from 60 kV/cm to 90 kV/cm for 1% strain-compensated GaInAs/AlInAs structures targeted around 4.3 µm. The laser frequency is also constrained to a window in order to avoid leakage of some part of the individuals to lower transition energies. An important choice is that we do not vary the number of layers in the structures. This supplementary variation would require a large empirical knowledge on the optimization method and probably would require a larger population by generation. However this type of algorithm is closer to the ideal genetic algorithm. Another potential problem or ill-defined variable is the dopant profile of the structure. During the optimization process the doped layers remain the same as well as the volume concentration. In a future revision of the method, it would be important to normalize each structures to the same sheet carrier density.

The random variation is constrained in the range of 20% of the initial thicknesses: we have found this value very empirically. If we increase this value in order to explore a larger portion of the phase space the population needs to be increased very strongly. If we keep constant the number of individuals, the merit function is not optimized. If we decrease the variation to 5% by thinking that the initial structure is near an optimum, we obtain basically no optimization even if we compute a large number of generations (approx. 40). However it is very hard to find rules for these parameters and we only report the empiric knowledge we have with the tested structures. Nevertheless for about 20% of variation, we clearly see an optimization in a few generations (approx. 8).

For each generation we keep a best pool of 32 individuals and we perform the random variations on these ones in order to build the next generation. We have also tried to vary this parameter, but the required number of generations seems to increase rapidly as we increase the number of individuals in the best pool.

On the side of computation, we basically submit to a cluster\textsuperscript{7} all the computations for each individuals in a generation at once. More precisely we divide the generations in chunks of 128 individuals in order to avoid to waste the computational time with jobs that take about 3 minutes per individual. We have coded scripts that automatically check if the generation needs to be updated (if all the jobs are done: we define a wall-time to avoid infinite self-consistent loops on pathologic structures) and we therefore let run the optimization process by itself. A typical optimization takes about 6 hours to 48 hours, depending on the structure to be optimized.

However an important part of the algorithm is missing at the present: we cannot test the stability of an optimized structure. Although we have some estimators like the distribution of the merit function as a function of the generation (some perfect structures can be

\textsuperscript{7}ETH Cluster Brutus: brutus.ethz.ch
therefore avoided), a stability test is missing. For the latter we would have to vary very locally the thicknesses of the layers and compute the corresponding standard-deviation of the merit function. A narrow distribution clearly indicates a stable structure while a broader one may indicate an unstable optimum. At the end we should be able to plot the variation of the merit function against the average distance from the optimized structure: the spreading along the curve would give a clear indication on the stability.

We have also to mention an important bias in the optimization algorithm. As we need to choose the coupling barriers for a given structure, we first have chosen to keep the same barriers all along the generations. This yields very pathological structures where the instabilities found in pure incoherent scattering models reappear in the sub-period of the structures. We have therefore imposed the injection barrier only with a fixed size and we have chosen one random coupling barrier inside the period. We have also imposed constrained on the magnitude of the coupling energies. This method have given pretty much better results and the main pathologies of the model are avoided. However we are clearly limiting the discovery of new structures as we have to increase drastically the number of individuals in order to decrease the impact of the randomness of the intra-period coupling barrier. Here it is absolutely sure that a model which dynamically choose the states in resonance and apply coherent/incoherent model when needed would give new results. In the very last version of the algorithm we however choose the coupling barrier inside the period by computing a merit function. The latter selects the barrier that makes the Fermi’s Golden rule as accurate as possible inside the sub-periods. The merit function on the coupling barrier is however very rough: it computes the energy differences between all pair of states in the sub-periods. The coupling barrier that maximizes this merit function is then selected. In particular this avoid weak anti-crossings with a large separation inside the sub-periods.

Some results are shown in Fig.(8.27) for a 1 % strain-compensated mid-infrared structure published in [117]. The original layer sequence is given by: \( \text{3.5/1.3/ 1.4/4.1/ 1.7/3.7/ 2.5/2.8/ 1.5/2.6/ 1.6/2.4/ 1.7/2.2/ 1.9/2.1/ 2.1/2.0/ 2.3/1.8/ 2.4/1.8, all thicknesses are in nm, boldfaced thicknesses indicated barriers in Al}_{0.665}\text{In}_{0.335}\text{As while roman thicknesses indicates wells in Ga}_{0.365}\text{In}_{0.635}\text{As. The underlined layers are Si doped with a volume concentration of 2 \cdot 10^{17} \text{ cm}^{-3}. The sheet carrier density is } N_s = 2.06 \cdot 10^{10} \text{ cm}^{-2}. \text{ The period length is } L_p = 49.3 \text{ nm.} \)

In Fig.(8.27a) the merit function is shown for individuals as a function of the generation. The optimization process shows a clear increase of the merit function after the 10\textsuperscript{th} generation. The slope is still positive afterwards but it seems that an optimum has been reached. The number of lasing structures compared to the total population in a generation is not
Figure 8.27: Automatic optimization of a mid-infrared 1% strain-compensated structure.
(a) The best individuals are shown as a function of the generation. The symbol $i\#n$ represents the merit function of individual number $n$ as a function of the generation. All the individuals were sorted, the first ones with the higher merit function values. (b) The distribution inside a generation is shown as a function of the individual number. The symbols $g\#n$ represents the curve for generation number $n$. In the inset the curves are limited to the best pool on which the next generation is based. The graph shows a clear optimization for all the individuals in the best pool as a function of the generation. More details are given in the text.
very high as for the individual #1024 the merit function is zero. The lasing structures roughly represent 3.1 % of the total population.

In Fig.(8.27b) the distribution is computed up to the 14th generation as a function of the individual number. This gives the distribution of structures inside a generation. We observe that there are a few excellent structures and the curves drop of about 20 % directly afterwards. However if we consider the inset where we have zoomed on the structures that are retained for the next generations (best pool), the merit function is clearly optimized, as it can be seen also shown in Fig.(8.27a).

The important question is which optimized structure should be chosen and tested experimentally. The stability curve is here missing and we have to develop some heuristic arguments. For sure we do not want to select the structures with the highest merit values as we are pretty sure that they will show instabilities. If we consider the curves in Fig.(8.27) we have to select a structure basically at the average position in the best pool (see inset). The latter does not give the best performances but is probably stable as many structures seem to be equivalent in this region.

In Fig.(8.28) we show the band-structure of the optimized structure. The sheet carrier density has been normalized to $N_s \approx 1.48 \cdot 10^{10}$ cm$^{-2}$. The strain between well and barrier material was also increased. The optimization was made for Ga$_{0.365}$In$_{0.635}$As/Al$_{0.665}$In$_{0.335}$As, but the structure was grown in Ga$_{0.275}$In$_{0.725}$As/Al$_{0.724}$In$_{0.276}$As. The optimized structure operates at a lower electric field, optimizing the electrical power. The injector region is also optimized.

For the optimized structure, the maximal current corresponds to the alignment between the first excited-state of the injector region and the upper laser-state. This feature was systematically observed in computer optimized structures. The injector ground-state plays the role of a charge reservoir but the injection is achieved with excited-states. This design reduces the field-strength required for injection alignment while the depopulation of the active subbands and the thermalization of electrons is kept as efficient as possible.

In the optimized structure, the width of the injector miniband is made thinner and the coupling between the states is increased.

The depopulation of the active subbands is also made more efficient by a better matching of the extraction resonances and an increased coupling strength. However, as the gain is not accurately predicted for thin injector barriers (8.4.1), the thickness of the injector barrier is not optimized (fixed to a constant value).

We have simulated the transport in order to compute the gain in the optimized structure with a uniform electron temperature for a lattice temperature of 300 K. The computation
EV1429
t = 300K

Figure 8.28: The band-structure of the optimized sample EV1429 is shown. The electric field roughly corresponds to the maximal current. Two periods are shown. The layer sequence is: 3.5/1.1/ 1.3/3.8/ 1.0/3.5/ 1.8/2.7/ 1.9/2.6/ 1.5/2.3/ 1.4/2.1/ 2.2/1.9/ 2.0/1.9/ 1.9/1.7/ 2.4/1.7. The material system is Ga$_{0.275}$In$_{0.725}$As/Al$_{0.724}$In$_{0.276}$As. The structure is Si doped in barriers to $1.0 \cdot 10^{17}$ cm$^{-3}$ and in wells to $1.81 \cdot 10^{17}$ cm$^{-3}$, yielding a sheet carrier density of $N_s \approx 1.48 \cdot 10^{10}$ cm$^{-2}$. The period length is $L_p = 452$, repeated 35 times.

converges with $T_e \approx 600$ K and shows a peak optical gain of 20 cm$^{-1}$. The intra-subband thermal distributions in the subbands are limiting the model when the photon energy is as large as 282 meV (corresponding to 4.4 µm).

Preliminary experimental validation of this optimization process are available at the time
of writing. However a more systematic study needs to be performed.

As the optimization concept and technique is promising, the algorithm describe above has
to be tested widely on various designs and for various merit functions. Nevertheless the
underlying structure of the algorithm is already working and we may expect that these
automatic optimization techniques will be used in order to obtain better structures for
many applications. We have nevertheless to remark that the optimization of the wallplug
efficiency is anyway a large added-value for any applications if the optimization process
does not alter the required features of the structures for a special application. We therefore
expect that some work will be invested in order to optimize structures under constraints.
As we can already fix genes of the structure (disabling mutation), a first test should be to
limit the optimization to a few (critical) layers.

These optimization techniques have a very large potential, but the first step should be to
define a reliable stability criterion.
Chapter 9

Conclusions and perspectives

The transport model we have developed all along this work has yielded interesting results. The agreement between predictions and experimental results first in terms of current-voltage curves [96] and in a second step in terms of light-current curves [94] was surprising since the model is only based on effective atomic states. The success of these predictions is fundamentally due to two reasons. The implementation of coherent effects at a few barriers inside the structure and the calculation of the current in the density-matrix formalism is the first. The second is due to the implementation of the second-order current between zero-dimensional states by computing effective coupling parameters and using an iterative computation. However the model has clearly shown its limits when we have tried to predict the second-order gain in the single-quantum-well structures. The thermal intra-subband distributions imposed for all subbands in the structure is a clear limitation of the model. The first improvement of the model would be to discretize the reciprocal space at the resolution of the LO-phonon energy and compute out-of-equilibrium electron distributions, as least for the active subbands.

The main part of this work concerns mid-infrared energies. In the far-infrared we have only developed a four-state model in the density-matrix formalism in order to model a particular structure based on resonant tunneling for the injection and the extraction of carriers. The roll-over in the light characteristic due to the misalignment of the extraction doublet has been clearly observed [83]. However we failed to predict the current-voltage curve of far-infrared structures with our effective transport model. The first main problem is the sharpness of the injection resonance. As we cannot compute the depasing time correctly, the current curve is unstable. We can assume that we do not have implemented all the relevant scattering sources. As the electron-electron scattering is missing we can expect that the dynamics of electrons in our model is too much trivial as the transport mainly occurs below the LO-phonon energy. We have however not intensively tested the
model on far-infrared structures. In a first step it should be interesting to implement a pure phase dephasing time \([14]\) in order to stabilize the current and see how the intraperiod dynamics is modeled. However as it is the case for short wavelength structures, the transport needs to be resolved in the reciprocal space. The problem is nevertheless more complicated than in the mid-infrared as we cannot sample the reciprocal space with a characteristic energy like the LO-phonon energy in the mid-infrared. Three-dimensional scattering models solved by Monte-Carlo techniques \([11, 13, 15, 21, 53, 65, 66]\) or quantum models \([61, 62, 64, 75, 119]\) are clearly more efficient in simulating far-infrared samples.

Another, complementary, way of stabilizing the current curves would be to consider inhomogeneous broadening. For interface roughness the latter is expected to dominate homogeneous broadening as the strength of the latter decreases in \(L^{-6}\), while the former decreases only in \(L^{-3}\), where \(L\) is the typical well width in the structure. The inclusion of inhomogeneous broadening in the current-voltage curve will require to first solve the transport model for a collection of islets in the plane of the heterostructure and then to average the corresponding voltage curves.

As a final result we have discussed and shown some motivating results in the automatic optimization of mid-infrared structures. It is clear that these techniques have a great potential. It is however clear that we have to sophisticate our model both on the fundamental side by adding more physics (better evaluation of the gain for thin barriers, transport partially resolved in k-space, additional scattering sources, etc.) and on the algorithm side by defining a reliable stability criterion.

A central problem in our transport model is the necessity of choosing the coupling barriers. This issue can be addressed by dynamically choosing the right model (coherent or incoherent) for each pair of subbands. In this model we however need to define the current density carefully. Another way of solving this issue is to consider a density matrix model where the scattering operators are basis invariant \([36]\). We will not recover the fundamental basis invariance of quantum models. In both cases we need to choose a representation in order to compute the self-energies for quantum models and the scattering rates for scattering models. However as the basis invariance is imposed ad hoc in \([36]\), the underlying approximations are harder to control than in quantum models. However this possibly enables to solve the problem of the choice of the coupling barriers. As the ingredients are basically the same in both models (dynamical choice of the model and invariance of the scattering operator) they should give comparable results.

The important question is what is purposed behind the simulation of transport in quantum cascade lasers. On the side of quantum models it appears clearly that up to the treatment of non-diagonal terms in the reciprocal space, the model is very successful and have largely
contributed to clarify the nature of the charge transport in the quantum cascade lasers, especially for far-infrared wavelengths. However at the present the computing time is very high if we think to use this model for optimization. Nevertheless such general models have a clear future in special applications that require a fundamental modeling: for example in the far-infrared where the situation has still to be clarified [73, 73, 75].

If we turn back to incoherent, pure scattering models with zero dimensional states [24, 45] it seems that the discussion is closed, although such models may still give interesting results for simple diagonal structures and can also be used in contrast with more sophisticated models in order to illustrate the improvements: what we have done basically in this work.

Regarding three-dimensional scattering models with sophisticated screening models [11, 52, 53, 68] they will also help to understand the far-infrared structures and to model properly the carrier-distributions in subbands. They should however be sophisticated to implement coherent effects as it was done previously [14]. The computational time is also very high when the interaction between an optical field is taken into account.

The model developed in this work lies between the incoherent scattering models, as it has finally the same mathematically structure, and a kind of hybrid models [14], where resonant tunneling is modeled with second-order terms without the resolution in the reciprocal space. It is clear that our model cannot be used, at least in the present form, for investigation of new quantum effects on the side of transport. As we have said in the introduction, we are keeping the minimal model that reproduces the experiment. It can however be used for computing other relevant quantities, like non-linear optical susceptibilities and can be use as a test-ground for modeling effects in larger systems.

Nevertheless, we have to make some efforts in order to enable the computation of the carrier-distributions in the reciprocal space and we have to find a way to recover partially a basis invariant model. However we think that our model is well positioned for computing optimization of existing structures as we have already obtained promising results.

We conclude this work by discussing the results for the second-order gain. As we have discussed in the text we have obtained evidences of such mechanisms in the mid-infrared [95], also observed by another group [79] in different samples with different techniques. We cannot say however that we have observed a Bloch oscillator as the latter is clearly defined in a superlattice. However we have observed the gain mechanisms on which quantum cascade lasers rely and the latter was shown to give the Bloch oscillator gain in a superlattice [115]. The robustness of this gain with temperature was demonstrated [39]. An important question is whether this gain can be used in order to increase the operation temperature of quantum cascade lasers, typically in the far-infrared. We may answer that it naturally exists in far-infrared quantum cascade lasers and may already plays an important role in the
operation of these devices. Another answer should be that the intra-subband scattering potential is so weak that second-order gain cannot be observed for this reason. Therefore adding controlled roughness may enable gain without a net population inversion. Here we want to refute this idea as if the broadening is small the peak-gain, ever without a net population inversion, is greater. The special vectors $\mathbf{q}_\pm$ does not depend formally on the magnitude of the scattering potential. The author’s opinion is that far-infrared lasers rely on second-order gain for currents and temperatures regimes where the population inversion is weak.
Appendix A

An example of super-self-consistent computations

In this section we consider the structure published in [99]. The band-structure is shown in Fig.(A.1a) for an electric field of $F = 67$ kV/cm that corresponds to a current of $\approx 1$ kA/cm$^2$ and for a thermal distribution of carriers at a temperature of $T = 80$ K.

The sheet carrier density is very high: $n_s = 4 \cdot 10^{11}$ cm$^{-2}$. Due to the thick extraction barrier (labelled B), we expect that a large electron concentration will remains localized far from the doped layers. This implies that the Hartree self-consistent potential cannot be approximated by a thermal distribution in the whole period. The self-consistent potential needs to be computed from the electron concentrations in each subbands under transport regime. This will modify the alignments in the structure and necessarily the optical gain.

In Fig.(A.1a) the band-structure is shown with a thermal distribution in the period. As the charge resides mainly in the ground-state and since the wavefunction of the latter is not far from the dopants, the self-consistent potential can be roughly neglected.

However for this structure this is a bad approximation of the actual potential in the structure in a transport regime. As shown in Fig.(A.1b) the super-self-consistent$^1$ modifies largely the band-structure. In order to compute this self-potential we had to solve the transport model developed in (7) in a self-consistency loop described in (4.5.6). We have implemented all the non-radiative scattering sources in (5) and the dephasing times were computed by using the Ando model in Eq.(5.10). The parameters for interface roughness scattering are $\Lambda = 9.0$ nm, $\Delta = 1.2$ Å and $\kappa = 1.5$ nm and the energy for the bulk LO-phonon is 32 meV.

---

$^1$This is the Hartree self-consistent potential computed with subband populations obtained in transport regime.
Figure A.1: The band-structure of the structure in [99] is shown for an electric field of 67 kV/cm corresponding to a current density of \( \approx 1 \text{ kA/cm}^2 \). The coupling barriers are in shaded green and labeled A, B and C. The doped layers are shown in pink. (a) The self-consistent potential is computed from a thermal distribution in the period. (b) The same structure where the self-consistent potential is determined by the transport model developed in (7).

The transport model used for these computations requires to select at least one barrier in the structure where resonant tunneling is implemented. As shown in Fig.(A.1) we have chosen three relevant barriers labelled A, B and C.

Due to their thicknesses and to the design of the structure it is relevant to model resonant tunneling at these barriers. We have tried to remove one of these barriers and therefore
replaced resonant tunneling by incoherent scattering. This has introduced large instabilities and the convergence on the self-consistent potential could not been achieved. These instabilities arise since we do not have model properly the transport at these barriers. The unphysical instabilities found in (6.3.3) being reflected on the self-consistent potential.

In Fig.(A.2) the results are shown. The current-voltage curve shown in In Fig.(A.2a) was computed for a lattice temperature of $T = 80$ K. The self-consistent potential was computed with the transport populations. The current densities are comparable with the typical magnitude of the measured ones [99].

In Fig.(A.2b) the optical gain is shown for various field strength. For low current densities, two transitions can be distinguished, the first near 140 meV and the second near 185 meV. However, as the current is increased, the peak gain of strongest (185 meV) transition increases and its linewidth is broaden. The low-energy transition is absorbed into the main transition.

It is worth to mention that we do not have simulated the laser optical field intensity in the structure. It would be interesting to compute the optical power for the low- and high-energy transitions first separately and then at the same time. This would however require to include in the model another photon-flux density $S'$ for the low-energy transition.

As the structure is very sensitive to the charge distribution in the period, the laser action may change completely the dynamic range of the gain and of the current density. It is common to observe experimentally that the frequency of the laser remains constant for the most of the current dynamic-range, the latter being given by the stimulated emission term which models the interaction of the electrons with the intensity of the laser field in the cavity.

We have also to mention that we are basically computing the optical gain inside the sub-periods defined by barriers A, B and C. The inter-period optical gain is computed by direct radiative transitions between localized states. However the tunneling-assisted contributions to the optical gain are neglected. The latter may be important for this structure. In order to evaluate properly these contributions, we have to implement the model of E. Dupont [25].

The direct computation of the optical gain without the super-self-consistent potential (assuming a thermal distribution in the period) also shows two resonances in the gain profile approximately at the same energies than for the super-self-consistent computations. However any calculations with straight potentials are not reliable, excepted if one shows that the stimulated emission terms redistribute the carriers inside the period so that the electrons screen the dopants.
Figure A.2: The current-density versus the electric field is shown together with the corresponding optical gain for various electric field strengths. (a) The optical gain is shown. The curves are shifted for clarity. The zero optical gain line is reported with dashed horizontal lines for each curve. The current-density and the electric-field strengths are reported on the IV curve with the same color-code.
In conclusion we have examined a structure where the self-consistent potential that arises from carrier distributions in a transport regime modifies completely the band-structure. The simulations show that the properties of the structure are deeply modified and that we cannot rely on calculations made with a simple thermal model.

A perfect numerical agreement with the measurements is however difficult to obtain as on the one side the gain model needs to be completed and as, on the other side, out-of-equilibrium intra-subband distributions need probably to be considered for current densities larger than 1.5 kA/cm$^2$. As mentioned previously, a complete simulation needs also to account for two photon-flux densities in the cavity.

As the complete modeling of this structure is interesting we have mentioned some calculations here. However it is beyond the scope of this work to implement missing parts of the transport model.
Appendix B

Dephasing times: comparisons between models

As discussed in (5.2.2) we used at the time of writing the body of this work an expression which overestimate the pure phase contribution to the dephasing time in Eq.(5.11).

Here, we have corrected this expression by implementing the expression in the Ando model in Eq.(5.10).

In Fig.(B.1) we have computed the LIV curves for sample N655 investigated in (8.4.3), for various lattice temperatures, by using the effective transport model developed in this work with the computation of the optical power.

In Fig.(B.1a) the curves are shown for the model with an overestimated dephasing time, while in Fig.(B.1b) the same curves are computed with the T. Ando model [2].

The difference between these dephasing times are translated for the current density as an increase of about 5 % for the maximal the maximal current in the Ando model. The optical power is also increased by approximatively the same percentage. The shape of the curves are however basically the same, although the injection resonance in the Ando model is more pronounced. The curvature of IV curves is the same and is basically a function of the temperature.

The improvement of the model on the dephasing time is important conceptually, but it does not changes the conclusions given in (9) and all along the text. The agreement with the measurements is still very good as the threshold currents are basically not affected and the current-voltage relation is the same for the full dynamic-range of the current.

In Fig.(B.2) we have performed a same kind of comparison for the structure of N258 sample. The latter is based on a single-quantum-well active region.
Appendix B. Dephasing times: comparisons between models

Figure B.1: The LIV characteristic of sample N655 (two-phonon design) are simulated. The electron temperature is determined with the kinetic energy balance model. The losses of the cavity are $\alpha_{\text{cav}} = 5.5 \text{ cm}^{-1}$. The optical power is computer for a ridge of width $10.1 \text{ } \mu \text{m}$ and a facet reflectivity of $R = 0.25$. The number of periods is $N_p = 35$.

(a) Computation performed with the overestimated dephasing time given by Eq.(5.11).
(b) Computation performed with the dephasing time of the Ando model in Eq.(5.10). Discussion is given in the text.
Figure B.2: The IV characteristic of sample N258 (single-quantum-well design) are simulated. The electron temperature is determined with the kinetic energy balance model. (a) Computation performed with the overestimated dephasing time given by Eq.(5.11). (b) Computation performed with the dephasing time of the Ando model in Eq.(5.10). Discussion is given in the text.
As for structures that show a strong population inversion the results are slightly affected by a longer dephasing time in the Ando model, it was important to show that structures with completely different designs, showing a weak population inversion, are also little affected. The comparison results in similar remarks than for the N655 structure. The current density is increased in the Ando model by about 5 %.

As the current resonances in the structure are more sharp since the dephasing time is increased, the alignment of the band-structure is more apparent in the Ando model.

If we compute the population inversion for the N258 structure, we observe a population inversion for the lowest temperature (15 K). The latter was not observed for overestimated dephasing times. However it is marginal and has basically no impact on the discussion (8.2.9), as it is clear that for this type of design, out-of-equilibrium intra-subband distributions are required if we intend to predict the second-order gain with more accuracy.
Appendix C

Steady-state solutions with the Laplace average

In this appendix, we apply the Laplace average on simple functions and differential equations and we review the main properties of this transform we use in the body of this work.

In (6.6.2) practical definitions are used in order to enhance the readability of the equations. We will not use them here as they may lead to confusion.

**Definition**  The Laplace average $\mathcal{L}$ of a function $f(t)$ for $t \in [0, \infty)$ is defined by:

$$\mathcal{L}[f](s) \equiv s \int_0^\infty dt \, e^{-st} f(t)$$

This transform is linear and defined for $s > 0$. It exists when the integral in the above definition exists. In order to demonstrate the main relations we have to assume that $e^{-st} f(t) \to 0$ when $t \to \infty$ for $s > 0$. The function $f(t)$ is therefore dominated by the exponential.

The Laplace average is related to the usual Laplace transform $\mathcal{L}$ by:

$$\mathcal{L}[f](s) = s \mathcal{L}[f](s)$$

**Steady-state solution**  The Laplace transform is used in order to obtain the steady-state (long-time limit) solutions to differential equations. The steady-state solution is obtained by performing the limit: $s \to 0^+$.

This limit needs however to be perform carefully as poles or ill-defined functions may appear.
Interesting results are obtained by using Laplace average when the differential equations are well-conditioned. Simple problems being expected, the application of the Laplace average and the calculation of the steady-state limit on raw dynamical equations usually yields zero or ill-defined results.

For example, it is standard to use the Laplace average in order to calculate the linear response of a system to an oscillating field at a given frequency. If the response function \( \varphi(t) \) of the system is at the same frequency, the Laplace average can be used in order to determine the amplitude of the response. This is achieved by applying the transform on the envelop function only, as shown below.

**Transformation of derivatives**  From the standard Laplace transform identities or by direct calculations, we apply the Laplace transform to differential operators \( d^n/dt^n \).

If we assuming that \( f \) and its derivative \( f^{(n)} \) of order \( n \), with respect to \( t \) are dominated by the exponential \( e^{-st} \) for \( s > 0 \), we have:

\[
\mathcal{L}[f^{(n)}](s) = s \mathcal{L}[f^{(n-1)}](s) - s f^{(n-1)}(0). \tag{C.3}
\]

The Laplace-average therefore reduces the order of the differential equation as the right-hand side of the equation only involves the \( f^{(n-1)} \) function. The initial condition at \( t = 0 \) appears as: \( s f^{(n-1)}(0) \).

This explains the additional multiplicative \( s \) in the definition of the Laplace average. When the steady-state limit is taken, the initial conditions disappear as \( s f^{(n-1)}(0) \to 0 \) when \( s \to 0^+ \). The steady-state solution does not depends on the initial conditions.

If the above transformation rule for derivatives is applied recursively, a system of differential equations becomes algebraic. In principle it can be solved and the steady-state limit is taken at the end. However for usual systems the steady-state limit has to be taken very carefully. It is standard to compute the steady-state limit on some parts of the solutions first, solve the system algebraically and then perform the limit on the final solution.

In the following paragraphs we consider simple examples on analytic functions.

**Exponential decay**  We first apply the Laplace average on a function \( \varphi(t) = n_0(1 - e^{-at}) \), with \( a > 0 \). The integral can be computed directly and yields:

\[
\mathcal{L}[\varphi](s) = n_0 \left(1 - \frac{s}{s + a}\right). \tag{C.4}
\]

If we take the limit \( s \to 0^+ \), we obtain \( \mathcal{L}[\varphi](s) = n_0 \) which is the expected (long-time) limit.
Oscillatory function  If now we consider a function with $a, \omega > 0$:

$$\varphi(t) = n_0(1 - e^{-at}) \sin \omega t$$  \hfill (C.5)

This is an oscillation with an exponential envelop $(1 - e^{-at})$. If we compute the Laplace average we have:

$$\mathcal{L}[\varphi](s) = \frac{n_0}{2i} \left( \frac{s}{s - i\omega} - \frac{s}{s + i\omega} \right) - \frac{n_0}{2i} \left( \frac{s}{s - i\omega + a} - \frac{s}{s + i\omega + a} \right)$$  \hfill (C.6)

The steady-state limit $s \to 0^+$ yields: $\mathcal{L}[\varphi](s) = 0$. And the average steady-state value is zero.

However we can factorize the function $\varphi(t)$ in an oscillatory part $\sin(\omega t)$ and an envelop function $(1 - e^{-at})$. If we take the Laplace average on the envelop function only, the steady-state limit yields: $\mathcal{L}[\varphi](s) = n_0$, which is typically the information we want to compute from a steady-state solution.

The above result holds for a general oscillatory function of period $T$. We consider the Fourier expansion:

$$\varphi(t) = \sum_n c_n e^{i n \frac{2\pi}{T} t} \quad \text{with} \quad c_n = \frac{1}{T} \int_0^T dt e^{-i n \frac{2\pi}{T} t} \varphi(t)$$  \hfill (C.7)

The Laplace average is given by:

$$\mathcal{L}[\varphi](s) = c_0 + \sum_{n \neq 0} c_n \frac{s}{s + \frac{n 2\pi}{T}}$$  \hfill (C.8)

The steady-state limit yields:

$$\mathcal{L}[\varphi](0^+) = c_0 \equiv \frac{1}{T} \int_0^T dt \varphi(t)$$  \hfill (C.9)

which is the average value of $\varphi(t)$ on the fundamental domain $[0, T]$.

Product of functions  Let $\varphi(t)$ be defined on $[0, \infty)$. If it can be decomposed into a product $\varphi(t) = \lambda(t) \phi(t)$, where $\phi(t)$ is a periodic function and $\lambda(t)$ is analytical, then:

$$\mathcal{L}[\varphi](0^+) = \phi_0 \mathcal{L}[\lambda](0^+),$$  \hfill (C.10)

where $\phi_0$ is the average value of $\phi(t)$ on its fundamental domain.

The proof is given below. We first expand $\phi(t)$ in a Fourier series of period $T$, and $\lambda(t)$ in a Taylor series at $t = 0$. We have:

$$\phi(t) = \sum_n \phi_n e^{i n \frac{2\pi}{T} t} \quad \lambda(t) = \sum_{n=0}^\infty \frac{\lambda^{(m)}(0)}{m!} t^m$$  \hfill (C.11)
The product $\varphi(t)$ reads:

$$\varphi(t) = \sum_n \sum_{m=0}^\infty \frac{\phi_n \lambda^{(m)}(0)}{m!} \theta_{nm}(t) \quad \text{where} \quad \theta_{nm}(t) = t^m e^{in \frac{2\pi}{T}}$$

(C.12)

The Laplace average of $\lambda(t)$ is given by the computation of the Laplace average of $\theta_{nm}(t)$, we have:

$$\mathcal{L}[\theta_{nm}](s) = s \int_0^\infty dt t^m e^{in \frac{2\pi}{T}} e^{-st} = m! \frac{s}{(s - \frac{m 2\pi}{T})^{m+1}}$$

(C.13)

where the result $\int_0^\infty dx e^{-ax} x^k = k! / a^{k+1}$ was used. We have for $\mathcal{L}[\varphi](s)$:

$$\mathcal{L}[\varphi](s) = \sum_n \sum_{m=0}^\infty \Phi_{nm}(s) \quad \text{with} \quad \Phi_{nm}(s) = \phi_n \lambda^{(m)}(0) \frac{s}{(s - \frac{m 2\pi}{T})^{m+1}}$$

(C.14)

We can rewrite $\mathcal{L}[\varphi](s)$ as a sum of two functions:

$$\mathcal{L}[\varphi](s) = \mu(s) + \zeta(s) \quad \text{where} \quad \mu(s) = \sum_{m=0}^\infty \Phi_{0m}(s) \quad \text{and} \quad \zeta(s) = \sum_{n \neq 0} \sum_{m=0}^\infty \Phi_{nm}(s)$$

(C.15)

We have: $\zeta(s) \to 0$ when $s \to 0$, as $\Phi_{nm}(s) \to 0$ for $n \neq 0$.

If we take the Laplace average of the Taylor expansion of $\lambda(t)$, we have:

$$\mathcal{L}[\lambda](s) = \sum_{m=0}^\infty \frac{\lambda^m(0)}{s^{m+1}}.$$

(C.16)

The comparison between $\mu(s)$ and the above equation enables to write:

$$\mathcal{L}[\varphi](s) = \phi_0 \mathcal{L}[\lambda](s).$$

(C.17)

By taking the limit $s \to 0^+$ we obtain the result.

However it is worth to point out that the convergence of the limit $s \to 0^+$ is not obvious in the Taylor expansion, and that the periodic function $\phi(t)$ needs to have a non-zero average value. If it is not the case the result is trivial. In the linear response regime this shows that direct averaging yields a trivial result.

A better approach clearly involves the envelop function of the response of the system, even if the system response is not at the excitation frequency. A good example is the calculation of the second-order gain where the Laplace average is used on an envelop operator in the density matrix formalism (8.2.4).
List of Tables

2.1  Kane model parameters for various III-V alloys  
2.2  The error in the effective two-band model is computed for various III-V alloys  
6.1  Scattering times computed for a two-phonon active region at the alignment 
      field of 48 kV/cm
# List of Figures

1.1 Transport regimes in superlattices ........................................... 4
1.2 Typical quantum cascade structure in the mid-infrared .................. 4

2.1 Triangular lattice .......................................................... 11
2.2 Types of solids: metals, semiconductors, insulators ....................... 13
2.3 Crystalline structure of III-V semiconductors .............................. 14
2.4 The band-structure of InAs ............................................... 15
2.5 Band diagram with $k \cdot p$ approximation ................................ 20
2.6 Solutions of the Kane model .............................................. 25
2.7 Plot of the energy-dependent mass ....................................... 28
2.8 Phase diagram of the lattice constant ................................... 33

3.1 A simple heterostructure ................................................... 36
3.2 Type I and type II heterostructures .................................... 37

4.1 Bound-states in a two quantum-well region .............................. 57
4.2 Bound-states and the continuum ......................................... 59
4.3 One period of a quantum cascade laser .................................. 61
4.4 Alignment electric field .................................................... 63
4.5 Tight-binding problem ...................................................... 66
4.6 Tight-binding between two quantum wells .............................. 73
4.7 Self-consistent potential example .................................... 80
4.8 Thermal distribution of carriers in the period .......................... 83
List of Figures

4.9 Thermal distribution in a quantum cascade structure .................. 84
5.1 Illustration of $\Gamma_{\text{inter}}$ and $\Gamma_{\text{intra}}$ .......................... 90
5.2 Effect of vertical correlation on interface roughness scattering ............ 102
5.3 Form-factors for the LO-phonon interaction ................................ 114
5.4 Various scattering times in strain-compensated materials .................... 115
6.1 Rate equations in the active region of a two-phonon design ................. 127
6.2 Time dependent simulations of rate equations ............................. 130
6.3 Periodicity of quantum cascade structures ................................ 133
6.4 Fundamental spectrum for wavefunctions with an infinite coherence length 135
6.5 Extended basis for sample S1548 ........................................... 141
6.6 Instabilities in the current-voltage curves of S1548 .......................... 143
6.7 Two quantum states coupled by a barrier ................................. 145
6.8 Numerical results for a two-state system pumped by an external current ... 153
6.9 Three periods of a model superlattice ..................................... 154
6.10 Current injection resonance in a two-phonon design ........................ 160
6.11 Density-matrix model for resonant injection and extraction ................ 164
6.12 Band-structure of far-infrared N471 sample ............................... 165
6.13 Effective transport time computed for N471 ................................ 167
6.14 Detuning and coupling energies as a function of the electric field for N471 sample .......................................................... 168
6.15 Roll-over in the population inversion ..................................... 170
6.16 Leakage current in N257 sample ........................................... 173
6.17 Population-inversion ratio with leakage current ............................ 175
6.18 Illustration of the second-order current density ............................ 185
6.19 Illustration of the $\sigma$ parameters ...................................... 187
6.20 The $\sigma$ parameter plotted as a function of temperature .................. 189
6.21 Illustration of the population term driving current between ground-states in a superlattice .................................................. 192
7.1 The coherent coupling between periods is illustrated for problems with one coupling barrier .......................................................... 209
7.2 Illustration of a periodic system with several coupling barriers .......... 213
7.3 Illustration of the kinetic energy balance between subbands ............... 219
7.4 Numerical solution for the electronic temperature .......................... 221
7.5 Structure details of N655 and N258 ............................................ 224
7.6 Effect of second-order contributions on tunneling between a pair of subbands 225
7.7 Voltage-current curves: the predictions are confronted to measurements . 226
8.1 Effect of gaussian broadening on a mid-infrared transition ................. 244
8.2 Non-parabolicity effects on the first-order gain curve ....................... 247
8.3 Gain due to non-parabolicity in the mid-infrared ........................... 249
8.4 Illustration of second-order gain mechanisms ............................... 253
8.5 The population inversion in second-order gain as a function of $\hbar \Delta - \hbar \omega$ . 256
8.6 Effects of non-parabolicity on second-order gain curves .................... 257
8.7 Second-order gain computations, different regimes .......................... 259
8.8 Effect of gaussian broadening on second-order gain curves ............... 261
8.9 Non-parabolicity effects computed in the configuration of sample N258 . 264
8.10 Illustration of gain and absorption with second-order gain mechanisms .... 266
8.11 Details for sample N258 ...................................................... 268
8.12 Details for sample N123 ...................................................... 269
8.13 Spectral modal gain measurements and simulations for samples N258 and N123 .......................................................... 270
8.14 LIV curves for N258 sample ................................................... 272
8.15 Transport simulations for N258 sample .................................... 275
8.16 Gain simulations for N258 sample ........................................... 276
8.17 Transport simulations for N257 sample .................................... 277
8.18 Gain simulations for N257 sample ........................................... 278
8.19 Spontaneous emission rate .................................................... 280
8.20 Effect of resonant tunneling on power: roll-over before maximal current . 286
8.21 LJV curves for N471 sample ............................................. 287
8.22 Luminescence measurements for sample N471 ...................... 288
8.23 Differential resistance measurements for sample N471 ............... 289
8.24 Simulation of LIV curves and the electronic temperature ............. 294
8.25 Population distributions as a function of the energy ................. 295
8.26 Threshold current versus lattice temperature ........................... 296
8.27 Individuals distributions in automatically optimized structures ....... 301
8.28 Band-structure of the optimized structure ............................. 303

A.1 Band-structure with a large super-self-consistent potential ............ 312
A.2 Current-voltage curve and gains with super-selfconsistent potential .... 314

B.1 N655: Effect of dephasing time ........................................... 318
B.2 N258: Effect of dephasing time ......................................... 319
Bibliography


Curriculum Vitae

Identity
Romain Terazzi
Born on April 21\textsuperscript{th} 1978, Geneva, Switzerland
Swiss, Italian
Married, 1 Child

Languages
Français mother tongue
English learned at school, practiced in scientific context
Deutsch learned at school but less used

Education
Autumn 2002 Master Degree in Physics, Université de Neuchâtel, Switzerland
Diploma thesis in High Energy Physics group with Prof. J.-P. Derendinger on gravitation, entitled “A little less symmetry for a bit more of gravitation”.

June 1997 Maturité type A, Literary, Bienne, Canton de Berne, Switzerland

Employment
Oct. 11 - present Employed at Alpes Lasers SA, Neuchâtel, Switzerland, for device modeling and industrial optimization.
Jul. 07 - Oct. 11 Continuation of PhD studies in the group of Prof. Faist, ETH, Zürich, Switzerland

May 03 - Jul. 07 PhD studies in Physics in the group of Prof. Faist, Université de Neuchâtel, Neuchâtel, Switzerland.

Jan. 03 - May 03 Training course in the group of Prof. Faist, Université de Neuchâtel, Neuchâtel, Switzerland.

Teaching experience

2004-2007 Teaching assistant of Modern Physics (for physicists) with Prof. J.-P. Derendinger and Prof. P. Aebi, Université de Neuchâtel, Neuchâtel, Switzerland

2004 Teaching assistant of Informatics (C/C++ for physicists, engineers) with Dr. Elham Firouzi, Université de Neuchâtel, Neuchâtel, Switzerland

Awards

2002 Prix Jean Landry, Université de Neuchâtel, Neuchâtel, Switzerland

Workshops

September 2010 International Quantum Cascade Lasers School & Workshop, Villa Finale, Firenze, Italy

26-27 June 2006 NCCR General assembly, Leysin, Switzerland

6-8 July 2003 NCCR General assembly, Chexbres, Switzerland

January 16, 2012
List of publications

Journals

2011  

2010  


2009  


2008  

2007

C. Walther, M. Fischer, G. Scalari, R. Terazzi, N. Hoyler and J. Faist, Quantum cascade lasers operating from 1.2 to 1.6 THz, Appl. Phys. Lett, Vol. 91, 131122 (2007)


2006


Conference papers

2011


2010

G. Scalari, M. I. Amanti, R. Terazzi, C. Walther, M. Beck, J. Faist, Broadband THz lasing from a photon-photon quantum cascade structure emitting from 2.8 to 4.1 THz, 2010 Conference on Lasers and Electro-optics (CLEO) and Quantum Electronics and Laser Science Conference (QELS), San Jose, CA, May 16-21 2010


2009


2008


2007


Acknowledgements

The first person I want to thank is my thesis supervisor Prof. Jérôme Faist. This was a true pleasure to work with him. I learned a lot from his subtle understanding of the science and his imagination. I have appreciated discussions we had in the office, first in Neuchâtel and then in Zürich, where we were trying to understand things in deep and in every details. There was a lot of enthusiasm and this was very motivating. I also want to thank him for his unconditional support in complicated life situations. His understanding and his sensitivity have helped my family to overcome very complicated health issues. He also learned me that researchers write papers. This was not obvious for me.

I want to thank my first co-examiner, Prof. Andreas Wacker. First of all, he read my thesis from the beginning to the end in very short delays. He pointed out important questions. The answers will have a significant impact on the future development of the quantum cascade laser simulator we have designed. During my PhD studies I have followed with great attention all his contributions to the field. I have tried to implement in my research work some important results demonstrated by his research group. In particular I always was appreciative of the elegance of the model he has developed. I met him in Cap Code (USA) during the ITQW’05 conference where he gave me a crash course on non-equilibrium Green’s function formalism by sharing a pizza. I will not forget this moment!

I want to thank Prof. Andreas Vaterlaus for his kindness during the oral examination and for all the work he made with Mrs Wildi in order to prepare the examination in very short delays. In this connection, I wish to thank the administrative staff for its efficiency. In particular I wish to thank Mrs Erna Hug, the secretary of the Quantum Optoelectronics Group, and Mrs Irene Wildi, the secretary of the department of Physics. Their effectiveness has allowed me to focus primarily on science.

On the scientific and personal side I want to thank especially Tobias Gresch for the work we did together on second-order gain. The measurements he made on the bloch-samples have been a test-ground for a large part of the transport model we have developed. I was also very pleased by his friendship.
I also want to thank Giacomo Scalari for his help in understanding far-infrared structures and the fruitful discussions we had on physics but also on completely unrelated things. I hope we will work together in the future.

I wish also to thank many of my colleagues, for scientific discussions or simply time spent together. In arbitrary order: Mattias Beck, Christophe Walther, Alfredo Bismuto, Andreas Hugi, Richard Maulini, Lorenzo Sirigu, Milan Fischer, Andreas Wittmann, Maria Amanti, Dimitri Yarekha, Yargo Bonetti, Marcel Graf, Stéphane Blaser, Thierry Aellen, Lubos Hvozdara, Marcella Giovannini, Nicolas Hoyler, Laurent Nevou, James Lloyd-Hughes, Fabrizio Castellano, Valeria Liverini, Borislav Hinkov, Arun Mohan, Markus Geiser.

I want thank the Swiss National Science Foundation and the National Center of Competence in Research, Quantum Photonics for having sponsored my PhD studies.

I wish to thank Dr. Antoine Müller, the CEO of Alpes Lasers SA for having given me the opportunity to continue to work on inter-subband systems in a motivating industrial environment.

I want to thank my parents Marie-Thérèse and Giorgio for giving me interested in science in a general way and for their support in human and financial ways. In my early age my mother was always speaking, among other things, about the understanding of natural phenomenons. I remember well a few molecules of central interest in pharmaceutics as she was freshly graduated from the University of Geneva in Pharmacy. My father has introduced me to computers and he has motived me to understand their fundamental operation. Later when I began Physics studies at the University of Neuchâtel, he was always behind me. He gave me a strong motivation and was always interested by the scientific work in general and by my personal achievements, regardless of his very time-consuming professional activity of surgeon.

I also want to thank my late Zio, Don Andrea, for his support and all the deep discussions we had on science, law, philosophy and religion. His vision of justice will always be inside me.

I want to thank Himali, my wife, for all the work she carries not only at home, but also mainly for the education of our daughter Lisa. Her commitment for enabling this work to be concluded, in a good situation for our little family, disregarding pernicious diseases, is above all praise. I have started to write sewlab when I met her and since that time, she made me dream again and she gave me the inspiration and the strength to go ahead despite of many problems, even if some are intractable.

Finally I want to thank my daughter Lisa for the illustration on the cover page. It reminds me a dipolar far-field which is of course very interesting.