Monte Carlo modeling of carrier transport in crystalline materials

Electronic transport in crystalline and organic semiconductors – Lab 2

Francesco Bertazzi, Michele Goano

April 26, 2020

Contents

1. Introduction

Premise: most of the footnotes of this document are explanations for expert/interested readers; some of them are digressions or historical ephemera. They (the footnotes) provide a more detailed look at stuff that the main text glosses over; they can be glanced at, or skipped altogether if the reader wants. The laboratory assignments are marked in blue. Supplemental material can be found in [1, 2, 3, 4].

The starting point of our discussion is the Boltzmann transport equation (BTE) – the first equation ever written to govern the time evolution of a probability – proposed in the late nineteenth century by Ludwig Boltzmann in his kinetic theory of gases and generalized in different fields of physics [5] including, of course, carrier transport in semiconductors, where it takes the form (see Appendix A)

$$\frac{\partial f}{\partial t} + \underbrace{\underline{v}\cdot\nabla_{\underline{r}} f}_{\text{diffusion}} + \underbrace{\frac{1}{\hbar}\underline{F}\cdot\nabla_{\underline{k}} f}_{\text{drift}} = \left.\frac{\partial f}{\partial t}\right|_{\text{coll.}} \tag{1.1}$$

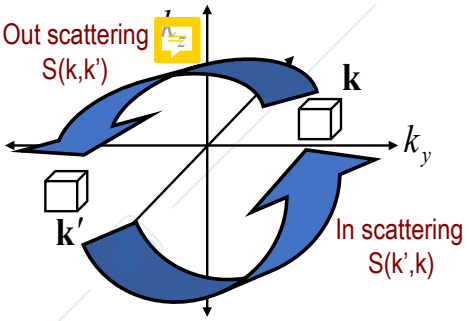
The relevant probability function here is the distribution $f(\underline{r},\underline{k},t)$ that describes the evolution in phase space of carriers treated as point-like particles with well defined geometrical position $\underline{r}$ and wavevector $\underline{k}$, subject to the Lorentz force

$$\underline{F} = \hbar\dot{\underline{k}} = -e\,(\underline{\mathscr{E}} + \underline{v}\times\underline{\mathscr{B}}) \tag{1.2}$$

and to scattering with photons, phonons, impurities, or other carriers, all described by the collision term in the right-hand side of (1.1)

$$\left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\text{coll.}} = \sum_{\underline{k}'}\left\{\underbrace{f(\underline{r},\underline{k}',t)S(\underline{k}',\underline{k})[1-f(\underline{r},\underline{k},t)]}_{\text{in scatt.}} - \underbrace{f(\underline{r},\underline{k},t)S(\underline{k},\underline{k}')[1-f(\underline{r},\underline{k}',t)]}_{\text{out scatt.}}\right\} \tag{1.3}$$

where $S(\underline{k},\underline{k}')$ is the rate at which a particle scatters from a state $\underline{k}$ to $\underline{k}'$



The first term in (1.3) represents the flow of particles into state $\underline{k}$, the second is the loss by scattering into another state. Each carrier population (electrons and holes) is described by its own distribution function, and the corresponding BTEs are coupled together (and to the crystal lattice) through the collision term. The description of the band structure is included in the BTE through the velocity term

$$\dot{\underline{r}} = \underline{v} = \nabla_{\underline{k}} E(\underline{k})/\hbar, \tag{1.4}$$

where $E(\underline{k})$ is the energy dispersion and the band index has been omitted to ease the notation.

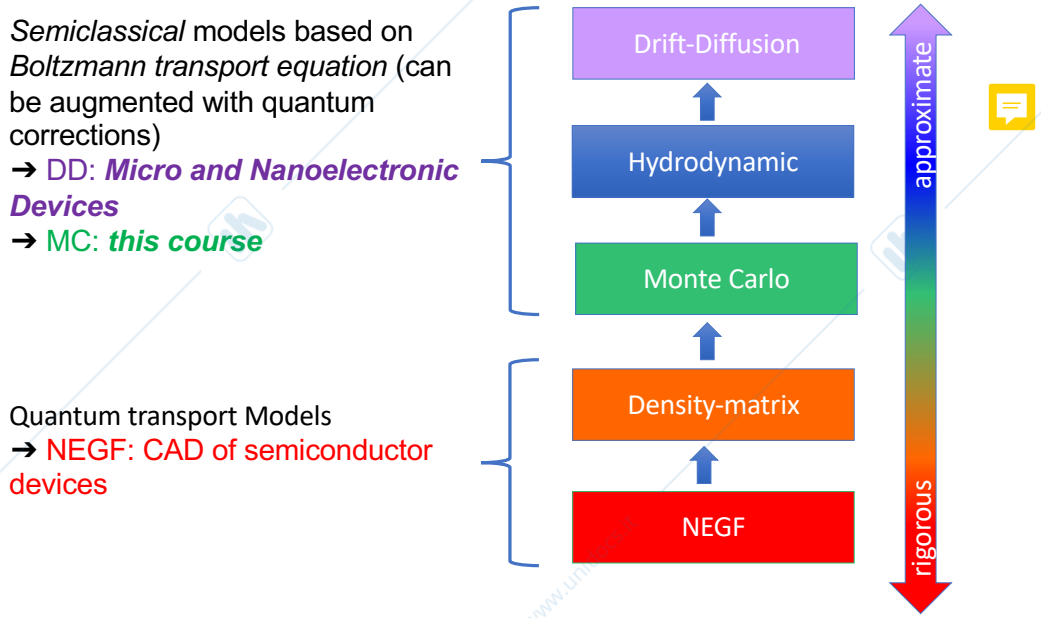


Figure 1.1.: Simulation tools for carrier transport in semiconductor devices, from rigorous to approximate.

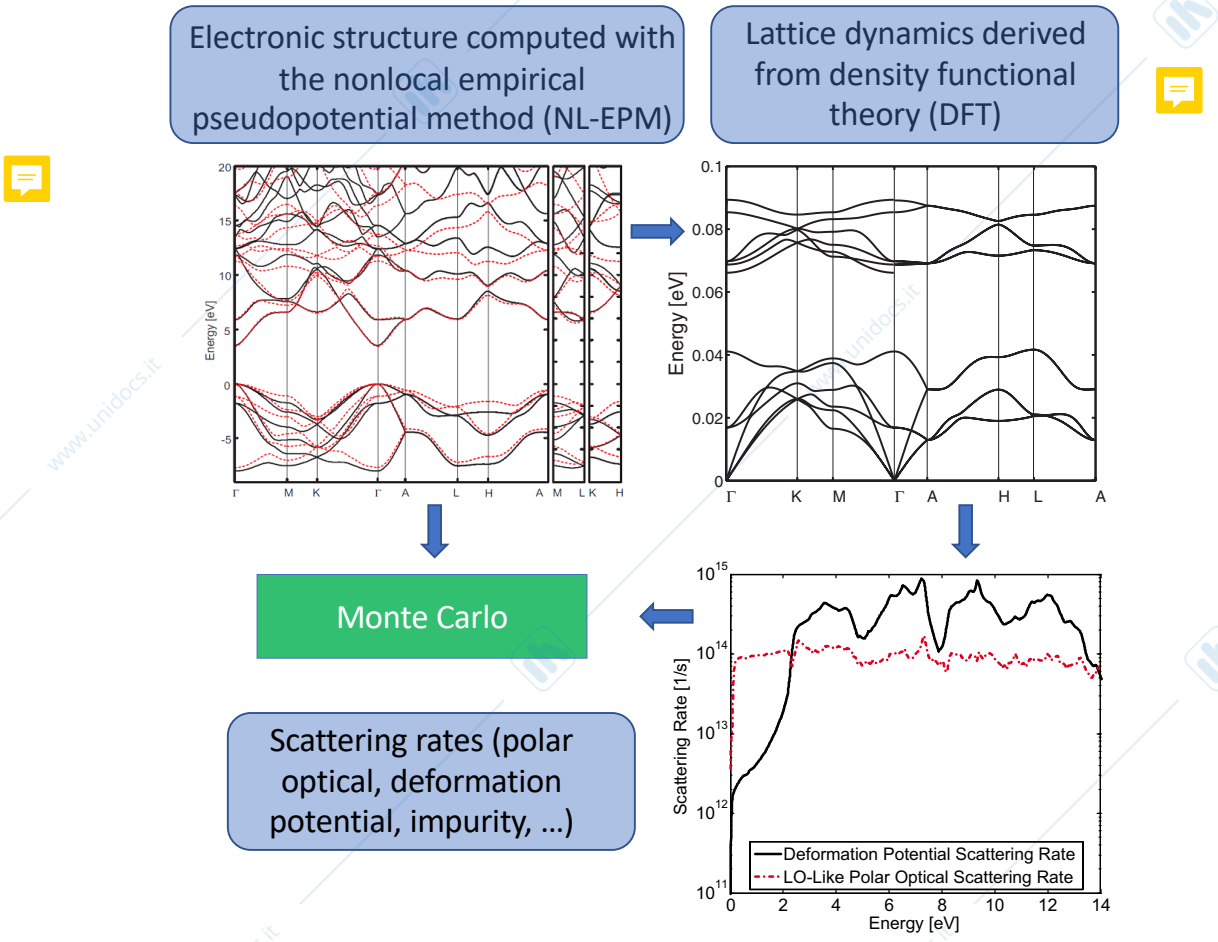


Figure 1.2.: Ingredients of a full-band Monte Carlo code.

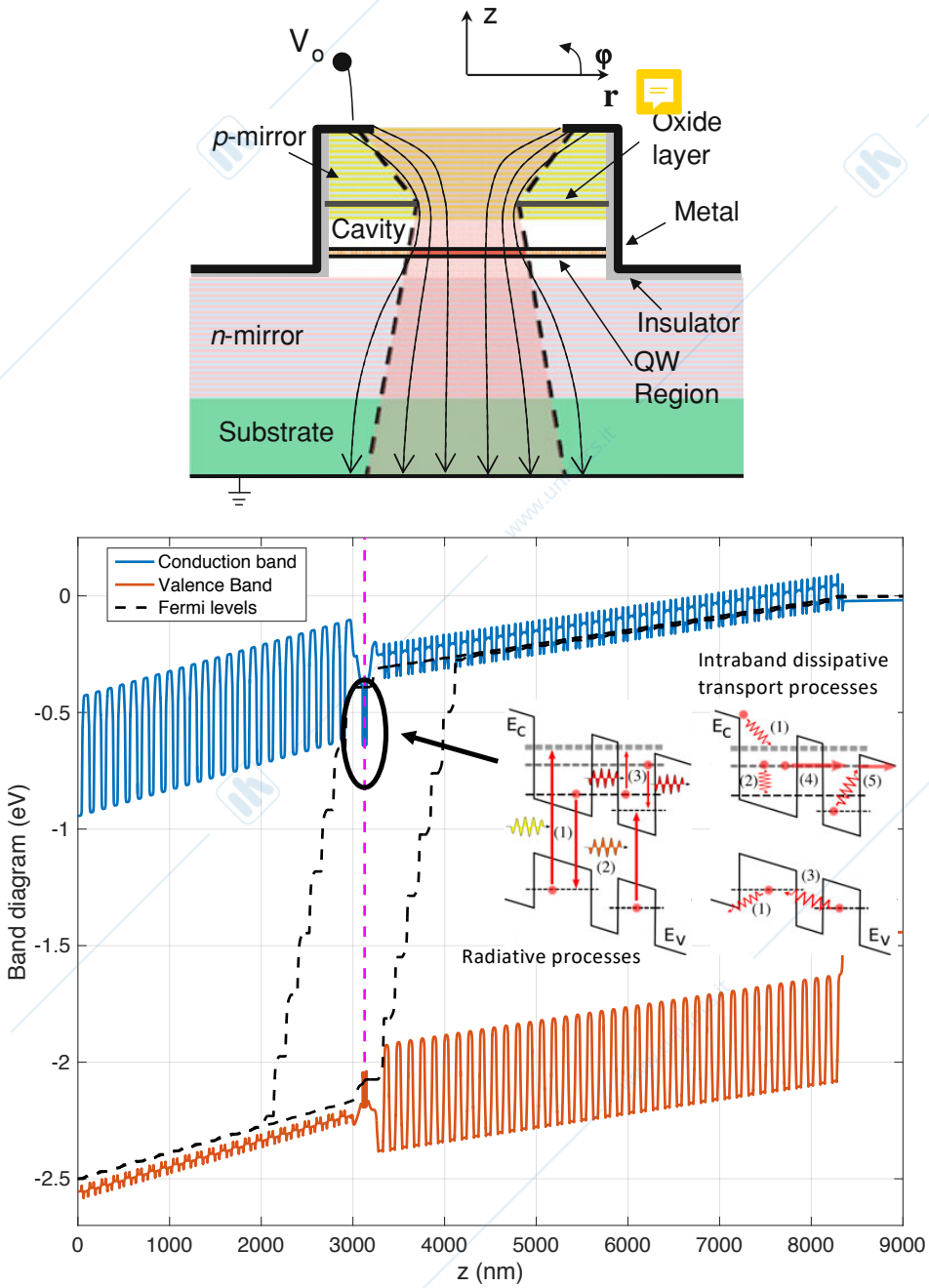


Figure 1.3.: Band diagram of a VCSEL. The relevant physical processes in the active region are: radiative transitions and coherent and dissipative quantum transport involving nonradiative intraband processes (see inset). The DBR mirrors must have a very high reflectivity, which means a large number of layers, typically tens of Bragg reflector pairs consisting of an alternating sequence of high and low refractive index layers with $\lambda/4$ thicknesses. Many effects, e.g., temperature profiles and current crowding make VCSELs three-dimensional structures. On the other hand, the active region of the VCSEL where the interesting physics happens is just a small fraction of the whole structure. The NEGF model requires a tremendous computational effort, and there is just no way that you can analyze the whole structure with a full quantum kinetic treatment.

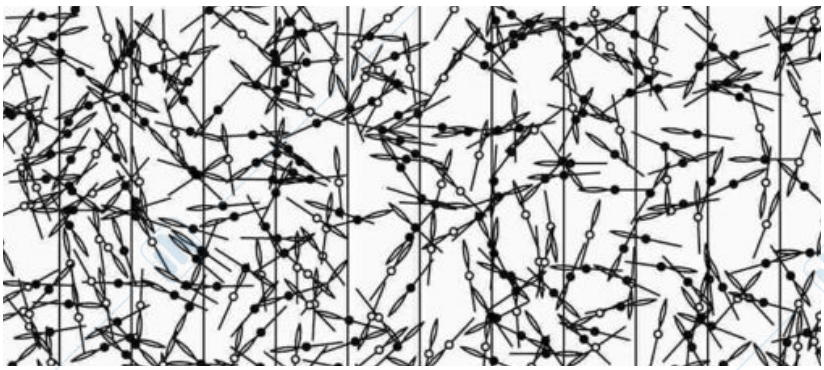


Figure 1.4.: XVIII-century Monte Carlo: needles are thrown onto the floor, separated by cracks. Needles are as long as the floor boards are wide, and they are dropped randomly, so that both the position and the angle are uniform random variables. The fraction of needles touching a crack is $2/\pi$.

The BTE is an integro-differential equation, nonlinear because the collision term depends on the distribution function $f(\underline{r},\underline{k},t)$ itself due to the Fermi statistical factors appearing in (1.3) (that also describe phase-space filling) or, in dealing with Coulomb interactions, via the Lindhard screening formula [4, Chapter 11]. Different approaches are available:

- Closed-form approximate solutions of the BTE may be obtained under homogeneous conditions and small electric fields, which allow to linearize the equation with respect to the field (relaxation-time approximation, see Section 4.1).
- Partial differental equations (PDEs) may be obtained from the BTE by expanding the unknown $f(\underline{r},\underline{k},t)$ into suitable basis functions (e.g., spherical harmonics [6]) and then solving for the expansion coefficients (expansion methods).
- Another possibility is to give up the goal to obtain the full distribution function, being satisfied with the knowledge of some ensemble averages of physical quantities, such as carrier density, velocity and energy (method of moments). Partial differental equations (PDEs) may be obtained taking the statistical moments of the BTE leading to an infinite hierarchy of equations, the lowest order approximation of which is the well known drift-diffusion model.
- A direct statistical solution of the BTE for the distribution function is provided by the Monte Carlo[1] method [7, 8, 2, 9], which consists in the simulation of the motion of an ensemble of carriers, subject to the action of electric and magnetic fields and of given scattering mechanisms, see Sections 4.2 and 4.3. The duration of the carrier free flights, i.e., the time between two successive collisions, and the states after scattering are determined stochastically according to the scattering probabilities, see Fig. 1.5. The books by Jacoboni, Reggiani, and Moglestue give full details about the method [7, 10, 8].

In introducing the theory of Monte Carlo carrier transport, it is useful to consider some of the basic hypotheses implicit in the semiclassical limit described by the BTE.[2]

[1] The earliest documented use of "random sampling" to solve an integral is that of Comte de Buffon who posed the problem: "Suppose that you drop a short needle on ruled paper – what is then the probability that the needle comes to lie in a position where it crosses one of the line?", see Fig. 1.4. A few years later, the idea was used by Laplace to estimate the value of π, which can be considered the first application of the Monte Carlo method. The origin of the word "Monte Carlo" dates back to one of the largest scientific collaborations ever undertaken, the Manhattan Project. As the backstory goes, the team lead by Enrico Fermi literally took their chances to solve a problem (the transport of neutrons through radiation shielding) that was too complex to be even described analytically. It was the first time that a method inspired by card games and gambling

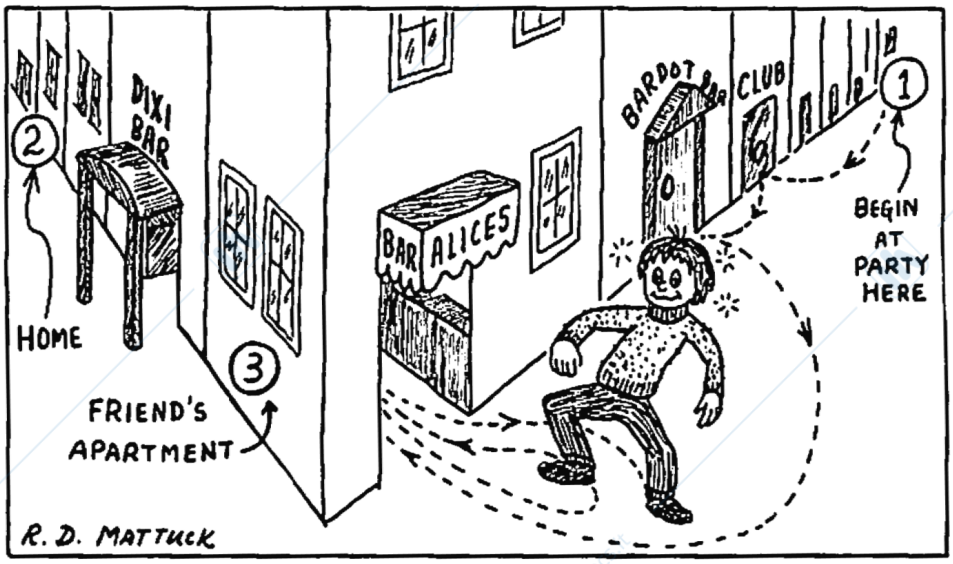


Figure 1.5.: Reproduced from [11].

1. The carriers have a well-defined position $\underline{r}$ and crystal momentum $\underline{k}$ as if they were classical particles. For this approximation to be acceptable, the characteristic length of the device L, say the channel length of a MOSFET, should be much larger than the size of the electron wave-packet given by the mean-free-path λ_ϕ (the average distance travelled by the electron between two collisions)

$$L \gg \lambda_\phi$$

and the energy scale of interest should be much larger than the uncertainty implied by the spread of the electron momentum.

2. The energy bands do not change too quickly in space, so that the concept of a local band structure may be accepted and level quantization can be neglected, see Fig. 1.6.

3. At any given point in $\underline{k}$-space, the energy bands are well separated in energy and the electric fields are sufficiently small so that we can describe the free-flight of the carriers between collisions with Newton's laws[3]

$$\dot{\underline{r}} = \underline{v}(\underline{k}) = \frac{1}{\hbar}\nabla_{\underline{k}} E(\underline{k})$$
$$\dot{\underline{k}} = \frac{\underline{F}}{\hbar} = -\frac{e}{\hbar}\left(\underline{\mathscr{E}} + \underline{v} \times \underline{\mathscr{B}}\right)$$

This approximation breaks down if, for some reason, the bands are close enough to allow band-to-band tunneling, see Fig. 1.7.

4. Boltzmann's Stosszahlansatz holds, i.e., all scattering processes are independent (no memory is conserved of where and when the previous collisions happened), local in space (they involve no

was employed for a serious scientific pursuit.

[2] N.B. that the BTE has resisted so far a rigorous derivation. The problem of deriving the macroscopic irreversibility that defines the "arrow of time" of the BTE from the microscopic reversibility of quantum mechanics eluded several attempts, including – and, obviously, starting with – Boltzmann himself [4, Chapter 18]. A heuristic derivation of the BTE obtained by means of conservation principles is reported in Appendix A.

[3] In wide band-gap materials, due to the small size of the Brillouin zone, multiple bands exist in relatively narrow ranges of energy and band-to-band tunnelling at band "kissing points" may be of importance, see Fig. 1.6b.

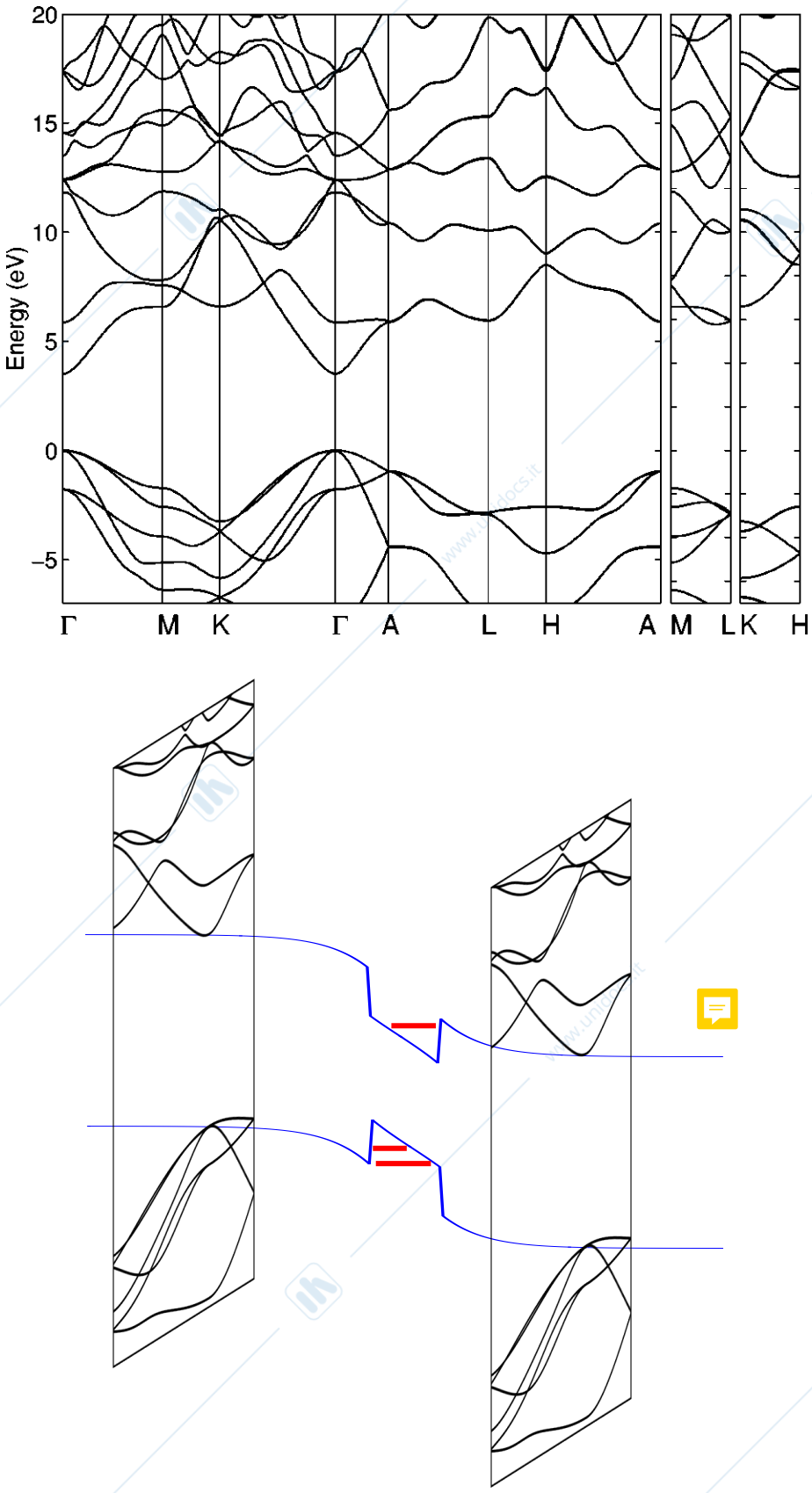


Figure 1.6.: Top panel: band structure of wurtzite phase GaN. Bottom panel: the local band structure approximation in the case of a GaN-based solar cell (a forward-biased *p-i-n* junction). The red lines represent confined levels; carrier confinement may be included in semiclassical approaches by means of quantum corrections.

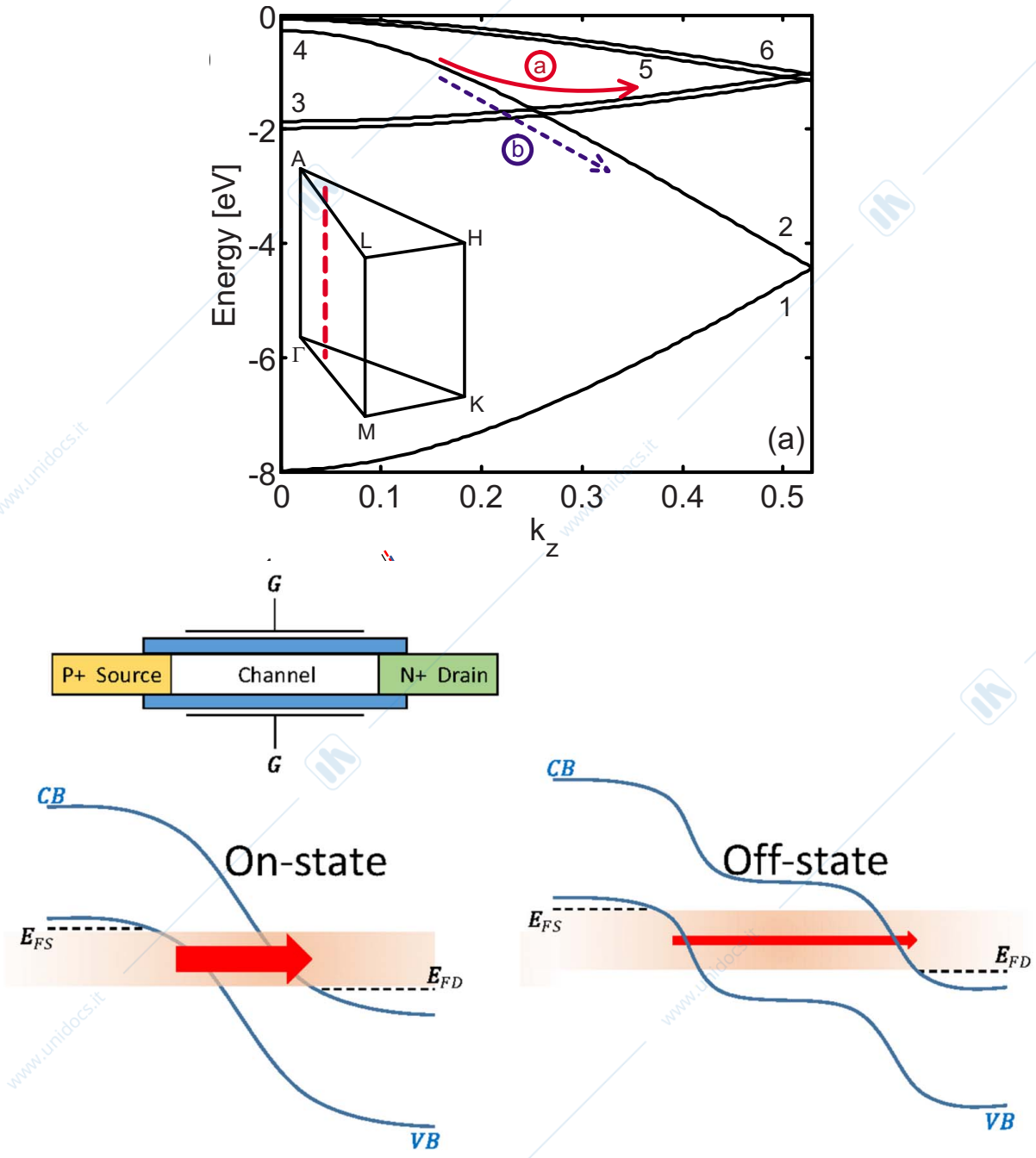


Figure 1.7.: Top panel. The six uppermost valence bands of bulk wurtzite GaN calculated along the direction in the Brillouin zone represented by the dashed line in the inset. If band-to-band transitions are not allowed, a ballistic hole in band four propagates coherently following the path indicated by the red arrow. This intraband tunnelling mechanism enables an additional (high-field) path indicated by the blue arrow. In the dissipative limit, when scattering mechanisms are present, the free flight propagation in momentum space would be interrupted by scattering events; in high field conditions the particle may travel a significant portion of the Brillouin zone, giving rise to the so-called *Bloch oscillations*. Bottom panel. Qualitative band diagram for a tunnel FET in the ON/OFF state. The shaded area indicates the tunneling window (where the conductions and the valence band overlap in energy) with the arrow representing electrons injected from the valence band of the source into the conduction band of the drain. As opposed to the intraband process considered above, the tunneling mechanism considered here is an interband process (connecting valence and conduction bands).

change of $\underline{r}$), instantaneous (their duration is negligible with respect to the free-flight time between successive collisions), and are perturbations weak enough to justify the use of the first-order (in perturbation theory) Fermi's Golden rule.[4]

5. An additional assumption, unrelated to the BTE derivation, is that time-domain transients are slow enough to justify the use of electrostatics rather than Maxwell's equations, meaning that the electric field $\underline{\mathscr{E}}$ may be determined only by the instantaneous distributions of the carriers via Poisson's equation, therefore neglecting propagation effects.

6. Carrier transport is described in terms of the electronic states of an infinite crystal, without any external potential applied (such as from the bias applied at contacts), and unperturbed by whatever scattering perturbation we shall consider shortly, as given by Bloch theorem, i.e., described by the product of a simple plane wave factor $e^{i\underline{k}\cdot\underline{r}}$ as for electrons of momentum $\hbar\underline{k}$ in free space, and a complicated fast-wiggling function $u_{\underline{k}}^{(n)}(\underline{r}) = u_{\underline{k}}^{(n)}(\underline{r}+\underline{R}_l)$ with the periodicity of the lattice ($\underline{R}_l$ is a lattice vector) that originates from the orbitals of the atoms that constitute the crystal

$$\psi_{\underline{k}}^{(n)}(\underline{r}) = \frac{1}{\sqrt{\Omega}} e^{i\underline{k}\cdot\underline{r}} u_{\underline{k}}^{(n)}(\underline{r}) = \frac{1}{\sqrt{\Omega}} e^{i\underline{k}\cdot\underline{r}} \sum_{\underline{G}} u_{\underline{G},\underline{k}}^{(n)} e^{i\underline{G}\cdot\underline{r}} \tag{1.5}$$

where Ω is the normalization volume, and n is the index labeling the band.[5] In addition to the symmetries in real space, there are also corresponding symmetries in the reciprocal space: the wavefunctions for values of $\underline{k}$ inside the first BZ give full knowledge of the electronic properties of the lattice.

[4] These assumptions break down in a few important cases: (i) very fast transients that have to be studied on a time scale comparable to the scattering rate [12]; (ii) high scattering rates, so that the initial state may decay appreciably by the time the scattering is completed (collision broadening) [12]; and (iii) high electric fields, which may transfer a significant amount of energy to the carriers during the finite time required for a collision, or, in other words, renormalize the states involved in the scattering process (intracollisional field effect [13]).

[5] The infinite field-free crystal is admittedly an over-idealized situation, and one may wonder if it makes even sense to consider it. In fact, the motion of the electrons in the crystal requires some electric field that drives the electrons. The reason why Bloch functions matter at all is the following [4, Chapter 12]. Assume that the length-scale over which the potential energy associated with the external field changes significantly is much larger than the distance between two collisions suffered on average by the electron, the so-called mean free path or coherence length. By changes significantly we mean a change comparable to the electron kinetic energy or to the thermal energy $k_B T$. This corresponds to situations in which all spatial variations induced by the external field are slow enough. In this case, one can ignore the spatial dependence of the electron wavefunction caused by the changing external potential, since coherent transport, characterized by a pure wavefunction, will not occur. Moreover, things do not change appreciably, as seen by the electron, between two collisions. Then, we can assume that locally, over the distance between successive collisions, the crystal appears almost field-free and we can approximate satisfactorily the electron wavefunction with the Bloch wave of the ideal crystal. This is the regime of semiclassical transport, since, besides the obvious role played by Quantum Mechanics in determining the band structure and Bloch wavefunctions, scattering occurs locally, ignoring interference effects between the wavefunctions scattered by other collisions that have occurred previously, and ignoring also the effect that the variation of the potential over the mean free path has on the wavefunction itself.

2. Lattice dynamics of crystals

Besides the electronic structure (see Lab. 1), one important ingredient needed to compute carrier-phonon interactions is the lattice dynamics of the crystal [14]. The physics of lattice vibrations is one of the best established branches of modern solid-state physics. There is a remarkable agreement between theoretical calculations and phonon dispersion curves measured by inelastic neutron scattering. Like electrons, phonons can also be represented as Bloch waves with energy (i.e., frequency) dispersion relations which give the values of the allowed energies versus Bloch wavevector. We will restrict the discussion to lattice vibrations in indefinitely extended crystals as perfect lattice periodicity will greatly simplify the analysis. In the non-relativistic limit, the Schrödinger equation to be solved for the systems of the interacting electrons and nuclei is (in SI units)

$$\hat{H}(\{\underline{r}_i\},\{\underline{R}_j\})\,\Psi(\{\underline{r}_i\},\{\underline{R}_j\}) = E\,\Psi(\{\underline{r}_i\},\{\underline{R}_j\}) \tag{2.1}$$

with

$$\hat{H}(\{\underline{r}_i\},\{\underline{R}_j\}) = \underbrace{\sum_i \frac{p_i^2}{2m} + \frac{1}{2}{\sum_{ii'}}' \frac{e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{r}_{i'}|}}_{\hat{\mathrm{H}}_\mathrm{e}} + \underbrace{\sum_j \frac{P_j^2}{2M_j} + \frac{1}{2}{\sum_{jj'}}' \frac{Z_j Z_{j'} e^2}{4\pi\varepsilon_0|\underline{R}_j - \underline{R}_{j'}|}}_{\hat{\mathrm{H}}_\mathrm{nuclei}} - \underbrace{\sum_{ij} \frac{Z_j e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{R}_j|}}_{\hat{\mathrm{H}}_\mathrm{e-nuclei}}, \tag{2.2}$$

where $\{\underline{r}_i\}$, $\{\underline{R}_j\}$ are the collections of electron and nuclear coordinates, respectively, and the total Hamiltonian includes the kinetic energy of electrons, the kinetic energy of the nuclei, and all possible electron-electron, electron-nucleus and nucleus-nucleus interactions. The factor $1/2$ in front of double sums is needed to avoid counting each pair twice. Z_j is the atomic number of the nucleus, $\underline{p}_i$ and $\underline{P}_j$ are the momentum operators of electrons and nuclei, respectively, and $\sum'$ means summation over pair of indices that are not identical. The eigenvalues E and the eigenfunctions $\Psi(\{\underline{r}_i\},\{\underline{R}_j\})$ of the combined electronic-nuclear system are called the *vibronic energies* and *vibronic wavefunctions*, respectively. Obviously, the many-particle Hamiltonian (2.2) cannot be solved without simplifications, since electronic motion is coupled to ionic motion.

We therefore invoke two approximations

1. First, we separate electrons into two groups: valence electrons and core electrons. The core electrons are those in the filled orbitals. These core electrons are mostly localized around the nuclei, so they can be lumped together with the nuclei to form the so-called *ion cores*. As a result of this approximation the indices j and j' will, from now on, denote the ions while the electron indices i and i' will label only the valence electrons.

2. The second approximation we invoke relies on the observation that, as the ions are much heavier than the electrons, their motion can be treated separately. Even if the ions move away from their equilibrium (lowest-energy) positions, the electrons adjust instantaneously to the new nuclear positions, in other words, the electrons follow the nuclear motion *adiabatically*. On the other hand, the ions cannot follow the electronic motion, what they see is only a time-average adiabatic electronic potential. Therefore, it is appropriate to disentangle the ionic and electronic motion (Born-Oppenheimer or adiabatic approximation).

The original Hamiltonian (2.2)

$$\hat{H}(\{\underline{r}_i\},\{\underline{R}_j\}) = \underbrace{\sum_i \frac{p_i^2}{2m} + \frac{1}{2}{\sum_{ii'}}' \frac{e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{r}_{i'}|}}_{\hat{\mathrm{H}}_\mathrm{e}} + \underbrace{\sum_j \frac{P_j^2}{2M_j} + \frac{1}{2}{\sum_{jj'}}' \frac{Z_j Z_{j'} e^2}{4\pi\varepsilon_0|\underline{R}_j - \underline{R}_{j'}|}}_{\hat{\mathrm{H}}_\mathrm{ions}} - \underbrace{\sum_{ij} \frac{Z_j e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{R}_j|}}_{\hat{\mathrm{H}}_{\mathrm{e-ions}}}, \tag{2.3}$$

can be written as

$$\hat{H} = \hat{H}_\mathrm{e}^\mathrm{eff} + \hat{H}_\mathrm{ions} + \hat{H}' \tag{2.4}$$

where

$$\hat{H}_\mathrm{e}^\mathrm{eff} = \underbrace{\sum_i \frac{p_i^2}{2m} + \frac{1}{2}{\sum_{ii'}}' \frac{e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{r}_{i'}|}}_{\hat{\mathrm{H}}_\mathrm{e}} - \sum_{ij} \frac{Z_j e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{R}_{j0}|}, \tag{2.5}$$

is the electronic adiabatic Hamiltonian with the ions frozen at their equilibrium positions $\{\underline{R}_{j0}\}$, and $\hat{H}'$ is the perturbation due to the displacement of the ions

$$\hat{H}' = -\sum_{ij} \frac{Z_j e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{R}_j|} + \sum_{ij} \frac{Z_j e^2}{4\pi\varepsilon_0|\underline{r}_i - \underline{R}_{j0}|}. \tag{2.6}$$

Neglecting the interaction $\hat{H}'$, we can decouple the full vibronic wavefunction into vibrational and electronic contributions

$$\Psi(\{\underline{r}_i\},\{\underline{R}_j\}) = \Psi_e(\{\underline{r}_i\};\{\underline{R}_j\})\,\Psi_\mathrm{ions}(\{\underline{R}_j\}), \tag{2.7}$$

where

$$\hat{H}_\mathrm{e}^\mathrm{eff}(\{\underline{r}_i\};\{\underline{R}_{j0}\})\,\Psi_e(\{\underline{r}_i\};\{\underline{R}_{j0}\}) = E_e(\{\underline{R}_{j0}\})\Psi_e(\{\underline{r}_i\};\{\underline{R}_{j0}\}) \tag{2.8}$$

$$\hat{H}_\mathrm{ions}^\mathrm{eff}(\{\underline{R}_j\})\,\Psi_\mathrm{ions}(\{\underline{R}_j\}) = E_\mathrm{ions}\,\Psi_\mathrm{ions}(\{\underline{R}_j\}) \tag{2.9}$$

in which have defined the effective ionic Hamiltonian

$$\hat{H}_\mathrm{ions}^\mathrm{eff}(\{\underline{R}_j\}) = \underbrace{\sum_j \frac{P_j^2}{2M_j}}_{\text{kinetic energy of the ions}} + \underbrace{\frac{1}{2}{\sum_{jj'}}' \frac{Z_j Z_{j'} e^2}{4\pi\varepsilon_0|\underline{R}_j - \underline{R}_{j'}|} + E_e(\{\underline{R}_j\})}_{\text{potential energy for the ionic motion}}. \tag{2.10}$$

Eq. (2.7) means that the ionic wavefunction $\Psi_\mathrm{ions}(\{\underline{R}_j\})$ is not quantum mechanically correlated to the electronic coordinates (each ion carries its own electronic cloud with itself, as the electrons follow the nuclear motion instantaneously). The electronic wavefunction $\Psi_e(\{\underline{r}_i\};\{\underline{R}_j\})$ depends on the ion coordinates parametrically, but not on the ions' wavefunctions, meaning that the electrons adjusting instantaneously to the positions of the ions. The variables $\{\underline{R}_{j0}\}$ appearing in $\hat{H}_\mathrm{e}^\mathrm{eff}$ are simply parameters (instead of quantum dynamic variables); thus the effective Hamiltonian $\hat{H}_\mathrm{e}^\mathrm{eff}$ belongs to the class of parameter-dependent operators (the semicolon is used to denote this parametric dependence).

In conclusion, the main ingredients of a very general first-principles approach to carrier transport are the following:

1. The system of light particles (the electrons) is preliminarily studied with the heavy particles (the ions) fixed in a given spatial configuration; then the total ground state energy of the electronic system $E_e(\{\underline{R}_j\})$ is obtained as a function of the (parametric) coordinates $\{\underline{R}_j\}$ by solving the Schrödinger equation (2.8)
$$\hat{H}_e^{\text{eff}}(\{\underline{r}_i\};\{\underline{R}_j\})\,\Psi_e(\{\underline{r}_i\};\{\underline{R}_j\}) = E_e(\{\underline{R}_j\})\Psi_e(\{\underline{r}_i\};\{\underline{R}_j\}). \tag{2.11}$$
2. The equilibrium positions of the ions $\{\underline{R}_{j0}\}$ are obtained by minimizing the potential energy for the ionic motion
$$V(\{\underline{R}_j\}) = \frac{1}{2}{\sum_{jj'}}' \frac{Z_j Z_{j'} e^2}{4\pi\varepsilon_0 |\underline{R}_j - \underline{R}_{j'}|} + E_e(\{\underline{R}_j\}). \tag{2.12}$$
3. The adiabatic Schrödinger equation for electrons (2.8)
$$\hat{H}_e^{\text{eff}}(\{\underline{r}_i\};\{\underline{R}_{j0}\})\,\Psi_e(\{\underline{r}_i\};\{\underline{R}_{j0}\}) = E_e(\{\underline{R}_{j0}\})\Psi_e(\{\underline{r}_i\};\{\underline{R}_{j0}\}), \tag{2.13}$$
gives the band structure of the crystal, see Lab. 1 with Prof. Michele Goano.
4. The phonon dispersion relation we are looking for is obtained by solving (2.9)
$$\hat{H}_{\text{ions}}^{\text{eff}}(\{\underline{R}_j\})\,\Psi_{\text{ions}}(\{\underline{R}_j\}) = E_{\text{ions}}\,\Psi_{\text{ions}}(\{\underline{R}_j\}). \tag{2.14}$$
5. The perturbation term $\hat{H}'$, expressing the interaction between the electron cloud and the lattice vibrations, has been ignored so far. We can include it at the scattering level by means of perturbation theory, see Section 3.1.
6. Include all this information, band structure, lattice dynamics and scattering mechanisms, in a bulk (Section 4.2) or device (Section 4.3) Monte Carlo carrier transport code.

Let's now turn our attention to ingredient 4 of the above list, i.e., the lattice dynamics. Consider, for the time being, a linear chain of atoms with masses M_n. Define u_n as the (longitudinal) displacement of the nth-ion from its equilibrium position R_{n0} at a particular time (for simplicity we assume that the atoms can vibrate only along the direction of the chain),
$$u_n = R_n - R_{n0} \tag{2.15}$$
and
$$p_n = M_n \dot{u}_n \tag{2.16}$$
as the corresponding conjugate momentum. Assuming that the forces acting on the ions depend instantaneously on the nuclear positions, i.e., taking the so called electrostatic limit ($c \to \infty$, see footnote 5), in which retardation effects due to the finite propagation of light are neglected, the Hamiltonian of the ions can be written as
$$H_{\text{ions}}^{\text{eff}}(\{u_n, p_n\}) = \frac{1}{2}\sum_n \frac{p_n^2}{M_n} + V(\{R_{n0} + u_n\}). \tag{2.17}$$
The first term in (2.17) represents the kinetic energy of the ions, and the second term is the potential energy of the ions interacting via the inter-ionic potential V (mediated by the electronic cloud). Assume that, at equilibrium, V has a minimum value $V(\{R_{n0}\})$ for some particular set of values $\{R_{n0}\}$. Assume

also that the displacements u_n of the ions away from their equilibrium positions are small enough to be considered a small perturbation. So, having made the approximation that the ions are close to equilibrium, we can expand the ionic potential around equilibrium

$$V(\{R_{n0}+u_n\}) = \underbrace{V(\{R_{n0}\})}_{\text{a constant}} + \underbrace{\sum_n u_n \left.\frac{\partial V}{\partial u_n}\right|_0}_{\text{zero at equilibrium}} + \underbrace{\frac{1}{2!}\sum_{nn'} \left.\frac{\partial^2 V}{\partial u_n \partial u_{n'}}\right|_0 u_n u_{n'}}_{\text{harmonic term}} + \underbrace{\frac{1}{3!}\sum_{nn'n''} \left.\frac{\partial^3 V}{\partial u_n \partial u_{n'} \partial u_{n''}}\right|_0 u_n u_{n'} u_{n''} + \dots}_{\text{anharmonic terms}} \tag{2.18}$$

where the subscript 0 indicates that the derivatives have been evaluated at the equilibrium positions $\{R_{n0}\}$. The first term is constant and is not important for dynamical calculations,[1] and the second term must vanish in the equilibrium configuration. Therefore the first term in the above expansion relevant to the vibration of the crystal is the quadratic term in the displacements. Keeping only this quadratic term (harmonic approximation) gives the classical ionic Hamiltonian[2] [4, Chapter 9]

$$H^{\text{eff}}_{\text{ions}}(\{u_n, p_n\}) = \frac{1}{2}\left(\sum_n \frac{p_n^2}{M_n} + \sum_{nn'} D_{nn'} u_n u_{n'}\right), \tag{2.19}$$

where

$$D_{nn'} = \left.\frac{\partial^2 V}{\partial u_n \partial u_{n'}}\right|_0 \tag{2.20}$$

is the matrix of the interatomic *forces constants* describing the interaction between the ions, i.e., the proportionality coefficients connecting the forces acting on the ions with the displacements suffered by the ions.[3] Combining Hamilton's dynamical equations

$$\dot{p}_n = -\frac{\partial H^{\text{eff}}_{\text{ions}}}{\partial u_n} \tag{2.21a}$$

$$\dot{u}_n = \frac{\partial H^{\text{eff}}_{\text{ions}}}{\partial p_n} = \frac{p_n}{M_n}, \tag{2.21b}$$

we obtain the classical equation of motion of the nth-ion of mass M_n at position $R_n + u_n$

$$M_n \ddot{u}_n = F_n(\{u_n\}), \tag{2.22}$$

under the force

$$F_n(\{u_n\}) = -\frac{\partial V}{\partial u_n} = -\sum_{n'} D_{nn'} u_{n'}. \tag{2.23}$$

[1] Setting $V(\{R_{n0}\}) = 0$ is just a rescaling of the zero-reference for energies.

[2] We will restrict our treatment of the lattice dynamics of semiconductors to the harmonic approximation only. The cubic and higher-order terms, containing products of three or more displacements, describe the weak interaction between the vibrational modes. The anharmonic terms are important to describe, for example, phonon-phonon scattering and phonon lifetime. The phonon-phonon interaction is responsible for relaxation of non-equilibrium LO phonons generated by hot electrons in semiconductors. The relaxation of non-equilibrium optical phonons occurs through their decay into a pair of longitudinal acoustic phonons: LO $\rightarrow$ LA1 + LA2. If the phonon lifetime is not short enough to maintain the phonon population at equilibrium, the distribution of phonons becomes an unknown of the problem, in the same way as the electron distribution. Since the emission and absorption of phonons by the electrons depend upon the phonons present in the sample, it is necessary to obtain electron and phonon distributions self consistently. The phenomenon is known as *hot-phonons*.

[3] The inter-ionic forces contain two contributions: the direct ionion interaction due to their Coulomb repulsion, and an indirect interaction mediated by the valence electrons. In fact, the motion of one ion causes the electron distribution to change and in turn this rearrangement produces a force on the neighboring ions.

The set of coupled equations (2.22) can be solved looking for solutions periodic in time of the form

$$u_n(t) = \xi e^{-i\omega t}. \tag{2.24}$$

Despite the powerful simplifications and approximations that we have made so far, namely the decoupling of electrons and ions (the adiabatic approximation) and the description of the ionic motion linearized around their equilibrium position (harmonic approximation), the large number of degrees of freedom still makes the problem intractable. The idea, already adopted in the calculation of the electronic structure, is to invoke the periodicity of the crystal.

2.0.1. Monoatomic linear lattice

Let's consider a chain of equal atoms with mass M, whose equilibrium separation is the lattice constant a, i.e., $R_n = na$. Considering the translational invariance of the system, which requires the force constants $D_{nn'}$ to depend on n and n' only through their difference $(n-n')$

$$D_{n+m,n'+m} = D_{nn'}, \tag{2.25}$$

we look for solutions that are periodic both in time and in space, of the form

$$u_n(t) = \xi_q e^{i(qna-\omega_q t)}. \tag{2.26}$$

Substitution into (2.22) gives

$$-\xi_q M\omega_q^2 = -\xi_q \sum_{n'} D_{nn'} e^{-iq(na-n'a)}, \tag{2.27}$$

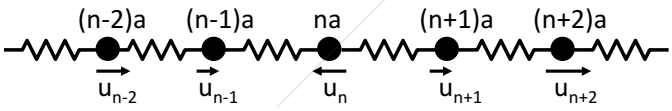
which can be written more compactly as

$$M\omega_q^2 = \mathscr{D}(q), \tag{2.28}$$

where

$$\mathscr{D}(q) = \sum_{n'} D_{nn'} e^{-iq(na-n'a)} \tag{2.29}$$

is the lattice Fourier trasform of the force constant matrix $D_{nn'}$, also known as the *dynamical matrix*. In order to proceed further, we need to make assumptions concerning the structure of $D_{nn'}$. Assuming that the atoms are connected by elastic springs with constant K and that the interactions are limited to nearest neighbors, i.e., only the nearest particles interact through the spring between them,



the effective ionic potential is

$$V(\{u_n\}) = \frac{K}{2}\sum_n (u_n - u_{n+1})^2 = \ldots + \frac{K}{2}(u_{n-1}-u_n)^2 + \frac{K}{2}(u_n - u_{n+1})^2 + \ldots, \tag{2.30}$$

with the corresponding equation of motion

$$M\ddot{u}_n = -K[(u_n - u_{n-1}) + (u_n - u_{n+1})]. \tag{2.31}$$

Looking for solutions of the type (2.26) we have

$$-M\omega_q^2 e^{iqna} = -K[2e^{iqna} - e^{iq(n-1)a} - e^{iq(n+1)a}], \tag{2.32}$$

or

$$-M\omega_q^2 = -K[2 - e^{-iqa} - e^{iqa}] = -4K\sin\left(\frac{qa}{2}\right), \tag{2.33}$$

which gives the dispersion relation

$$\omega_q = 2\sqrt{\frac{K}{M}}\left|\sin\left(\frac{qa}{2}\right)\right|, \tag{2.34}$$

see Fig. 2.1. Since (2.34) is periodic with period $2\pi/a$, it is sufficient to consider wavevectors between

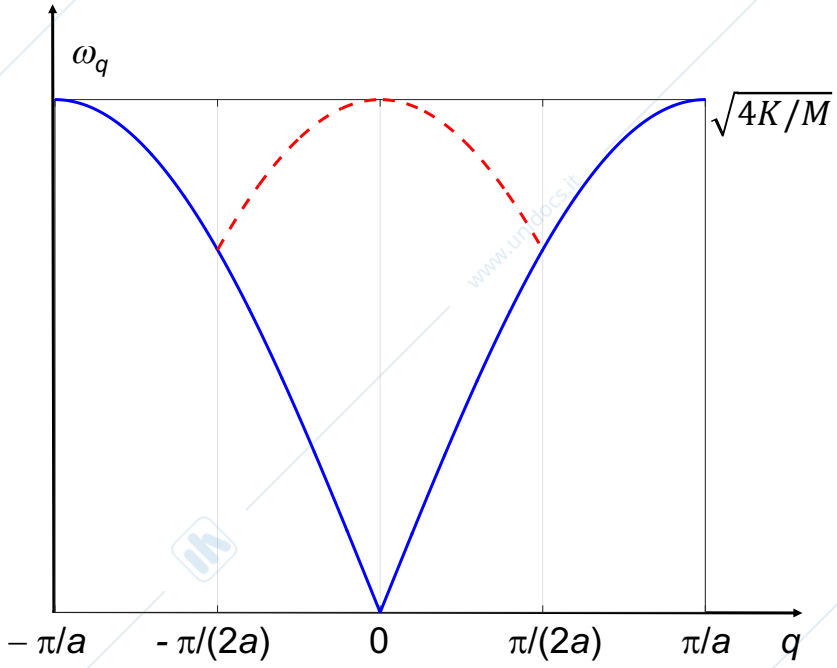


Figure 2.1.: Phonon dispersion curve (blue line) of a monoatomic linear lattice with nearest neighbor interactions and interatomic distance a. The Brillouin zone is the segment between $-\pi/a$ and $+\pi/a$. If we assume that the minimum translation vector is $2a$, the first Brillouin zone would be between $-\pi/(2a)$ and $\pi/(2a)$, and the acoustic branch outside of this region would be translated (or folded back) to $|q| \leq \pi/(2a)$. In other words, a new phonon branch (dashed red line) would appear, see the following discussion.

$-\pi/a$ and $+\pi/a$. This interval is the first unidimensional Brillouin zone (BZ). It is interesting to consider the normal modes in the long wavelength limit ($qa \ll 1$), when the wavelength is much longer than the interatomic distance, ω_q is approximately given by the linear relation, the phonon dispersion takes the form

$$\omega_q \approx v_l|q| \tag{2.35}$$

where

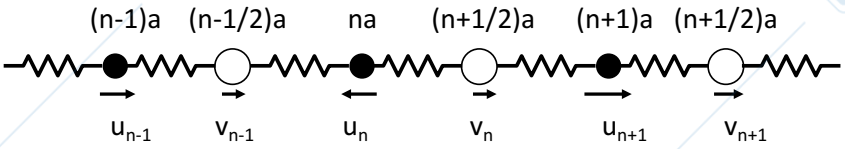
$$v_l = a\sqrt{K/M} \tag{2.36}$$

is the velocity of the longitudinal wave.

Notice that we have reduced the infinite set of equations of motion (2.22) to the homogenous problem (2.28) in the unknown amplitude ξ. This simplification is the direct consequence of the periodicity of the lattice expressed by (2.25).

2.0.2. Diatomic linear lattice

Since most semiconductors have more than one atom per primitive unit cell, we consider a linear chain composed of two types of atoms, with alternate masses M_1 and M_2 (diatomic linear chain). We assume that in the equilibrium configuration, the atoms with mass M_1 take the sublattice positions $R_n = na$, and the atoms of mass M_2 take the sublattice positions $R_{n+1/2} = (n+1/2)a$. We also consider the simple case of interactions between only nearest particles by means of elastic forces of spring constant K.



The results would not be essentially different for the more general case of different interatomic distances and spring constants. If u_n is the displacement of the atom with mass M_1 and v_n is the displacement of the atom with mass M_2, the classical equations of motion are[4]

$$M_1 \ddot{u}_n = -K(2u_n - v_{n-1} - v_n) \tag{2.37a}$$
$$M_2 \ddot{v}_n = -K(2v_n - u_n - u_{n+1}) \tag{2.37b}$$

for any integer number n. Considering the translational invariance of the lattice, we look for Bloch waves of the form

$$u_n(t) = \xi_1 e^{\mathrm{i}(qna - \omega t)} \tag{2.38a}$$
$$v_n(t) = \xi_2 e^{\mathrm{i}(qna + qa/2 - \omega t)} \tag{2.38b}$$

Replacing these expressions in (2.37b) we obtain homogeneous equations in the unknown amplitudes ξ_1 and ξ_2

$$-M_1\omega^2\xi_1 = -K(2\xi_1 - \xi_2 e^{-\mathrm{i}qa/2} - \xi_2 e^{\mathrm{i}qa/2}) \tag{2.39}$$
$$-M_2\omega^2\xi_2 = -K(2\xi_2 - \xi_1 e^{-\mathrm{i}qa/2} - \xi_1 e^{\mathrm{i}qa/2}), \tag{2.40}$$

or

$$(2K - M_1\omega^2)\xi_1 - 2K\cos(qa/2)\xi_2 = 0 \tag{2.41}$$
$$-2K\cos(qa/2)\xi_1 + (2K - M_1\omega^2)\xi_2 = 0. \tag{2.42}$$

A nontrivial solution is obtained if the determinant is zero, namely

$$\begin{vmatrix} 2K - M_1\omega^2 & -2K\cos(qa/2) \\ -2K\cos(qa/2) & 2K - M_1\omega^2 \end{vmatrix} = 0, \tag{2.43}$$

which gives the following eigenvalues (dispersion relation) [2, Ch. 5]

$$\omega_q^2 = K\left(\frac{1}{M_1} + \frac{1}{M_2}\right)\left[1 \mp \sqrt{1 - \frac{4M_1M_2}{(M_1+M_2)^2}\sin^2(qa/2)}\right]. \tag{2.44}$$

and the corresponding amplitudes satisfy the relation

$$\frac{\xi_1}{\xi_2} = \frac{2K\cos(qa/2)}{2K - M_1\omega_q^2}. \tag{2.45}$$

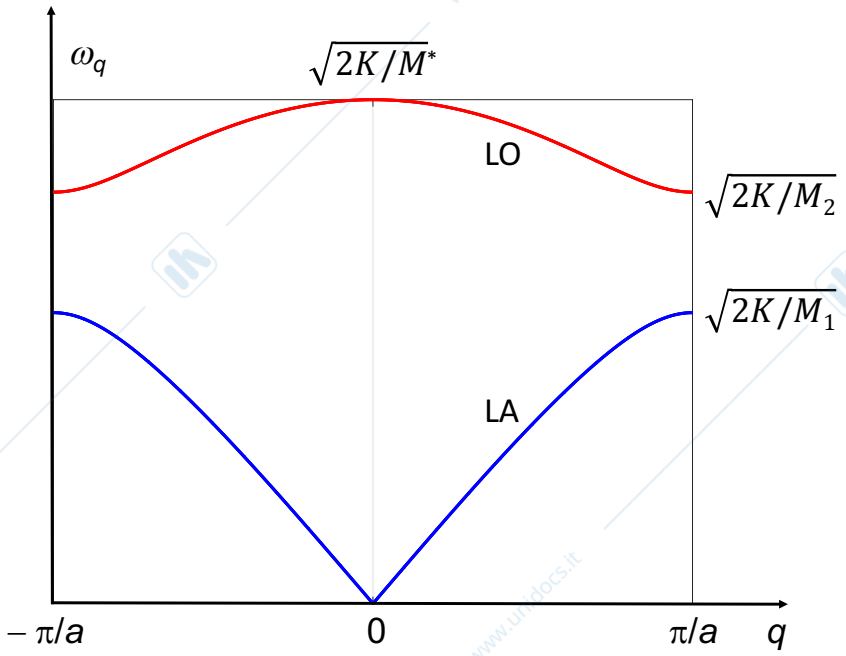


Figure 2.2.: Acoustic branch (blue line) and optical branch (red line) of the phonon dispersion relation for a diatomic linear lattice with nearest neighbor interactions. The Brillouin zone is the segment between $-\pi/a$ and $+\pi/a$. M_1 and M_2 are the atomic masses (in this case $M_1 = 2M_2$), M^* is the reduced mass. The longitudinal acoustic and optical modes and are labeled LA and LO, respectively.

Equation (2.44) shows that the dispersion relation of a a diatomic lattice has two branches, as shown in Fig. 2.2, the lower one called *acoustic* and the upper one called *optical*, with a frequency gap between them. If $M_1 = M_2$, the gap between the two branches disappears and we recover the result of the monoatomic linear lattice (the optical branch should be interpreted as the continuation of the acoustic branch in the doubled Brillouin zone). In the long wavelength limit ($qa \ll 1$), the dispersion relation of the acoustic branch and the corresponding amplitude can be approximated as

$$\omega_q \approx v_l |q| \tag{2.46}$$
$$\xi_1 \approx \xi_2, \tag{2.47}$$

where v_l is the longitudinal sound velocity

$$v_l = \frac{a}{2}\sqrt{\frac{K}{(M_1+M_2)/2}}. \tag{2.48}$$

Notice that (2.46) is the obvious counterpart of (2.35), with the average mass $(M_1+M_2)/2$ replacing M, while (2.47) means that the atoms vibrate in phase and with the same amplitude. As q approaches zero, the frequency of the optical branch tends to a constant value, and the amplitudes ξ_1 and ξ_2 have opposite signs and absolute values inversely proportional to the atomic masses

$$\omega_q \approx \sqrt{2K/M^*} \tag{2.49}$$
$$\xi_1 \approx -\frac{M_2}{M_1}\xi_2, \tag{2.50}$$

[4]Here we discuss in detail only the longitudinal modes, but it is obvious that both for the monatomic and the diatomic chain, for every wave vector and every branch, two (degenerate) transverse oscillations also exist. In this case, the displacements of the atoms are not along the chain but perpendicular to it.

in other words, the two atoms in the unit cell move in opposite directions, while the center of mass of the unit cell remains fixed (M^* is the reduced mass given by $1/M^* = 1/M_1 + 1/M_2$). If the two different atoms carry opposite charges, as a result of the redistribution of the charge associated with polar bonding, i.e., if the binding is at least partly ionic, then the oscillation is connected with an oscillating electric dipole. This allows it to couple to the electromagnetic light field, and this is why these oscillations are called *optical* modes.

The results so far show that for each atom in the basis (unit cell), one branch appears in the spectrum. We can extend this to three-dimensional crystals, where we find three branches for each atom in the unit cell. The number of acoustic branches is limited to three, as there can be only three ways in which the atoms vibrate in phase. So if we have more than one atom per unit cell, all but three of the modes must be optical, see next paragraph.

2.0.3. Dynamics of three-dimensional crystals

The lattice dynamics of real three-dimensional crystals is much more complicated, but the procedure that we will be using is the same. Labeling atoms with the compound index $n = (\gamma l)$, γ being the atom type and l the unit cell, the classical equations for the ions in the instantaneous positions $\underline{R}_{l\gamma} + \underline{u}_{l\gamma}$ under the forces $\underline{F} = -\partial V/\partial \underline{u}_{l\gamma}$ are

$$M_\gamma \ddot{u}_{l\gamma\alpha} = -\sum_{l'\gamma'\alpha'} D_{l\gamma\alpha,l'\gamma'\alpha'}\, u_{l'\gamma'\alpha'} \tag{2.51}$$

where $\alpha, \alpha' = x, y, z$. The force constants are still defined as the second derivatives of $V(\{\underline{u}_{l\gamma}\})$ evaluated at the equilibrium configuration

$$D_{l\gamma\alpha,l'\gamma'\alpha'} = \left(\frac{\partial^2 V}{\partial u_{l\gamma\alpha}\partial u_{l'\gamma'\alpha'}}\right)_0. \tag{2.52}$$

Translational invariance implies

$$D_{l\gamma\alpha,l'\gamma'\alpha'} = D_{m\gamma\alpha,m'\gamma'\alpha'} \tag{2.53}$$

if $l - l' = m - m'$, which suggests to look for solutions of the form

$$\underline{u}_{l\gamma}(t) = \underline{\xi}_{\underline{q}\gamma} e^{\mathrm{i}(\underline{q}\cdot\underline{R}_{l\gamma} - \omega_{\underline{q}} t)}. \tag{2.54}$$

Here, the polarization vector $\underline{\xi}_{\underline{q}\gamma}$ gives the the amplitude and the direction of the displacement of atom γ associated with wavevector $\underline{q}$. Replacing (2.54) in (2.51) gives the secular equation

$$\sum_{\gamma'\alpha'} [\mathscr{D}_{\gamma\alpha,\gamma'\alpha'} - M_\gamma \omega_{\underline{q}}^2 \delta_{\alpha\alpha'}\delta_{\gamma\gamma'}]\xi_{\underline{q}\gamma'\alpha'} = 0 \tag{2.55}$$

where we have defined the dynamical matrix of the crystal in reciprocal space

$$\mathscr{D}_{\gamma\alpha,\gamma'\alpha'}(\underline{q}) = \sum_{l'} D_{l\gamma\alpha,l'\gamma'\alpha'}\, e^{-\mathrm{i}\underline{q}\cdot(\underline{R}_{l\gamma} - \underline{R}_{l'\gamma'})}. \tag{2.56}$$

The condition that the homogenous system of equation (2.56) has a nontrivial solution requires the determinant to vanish

$$|\mathscr{D}_{\gamma\alpha,\gamma'\alpha'}(\underline{q}) - M_\gamma \omega_{\underline{q}}^2 \delta_{\alpha\alpha'}\delta_{\gamma\gamma'}| = 0, \tag{2.57}$$

which gives the vibrational frequencies $\omega_{\underline{q}}$. The dynamical matrix at a given $\underline{q}$ has size $3N_\gamma$, N_γ being the number of ions in the primitive cell. Taking the long wavelength limit, it is possible to show that of the total $3N_\gamma$ phonon modes, three are acoustic, and $3(N_\gamma - 1)$ are optical modes. A mode is called *longitudinal* (*transverse*) if the polarization vector $\underline{\xi}_{\underline{q}\eta}$ is parallel (perpendicular) to $\underline{q}$. Modes that involve

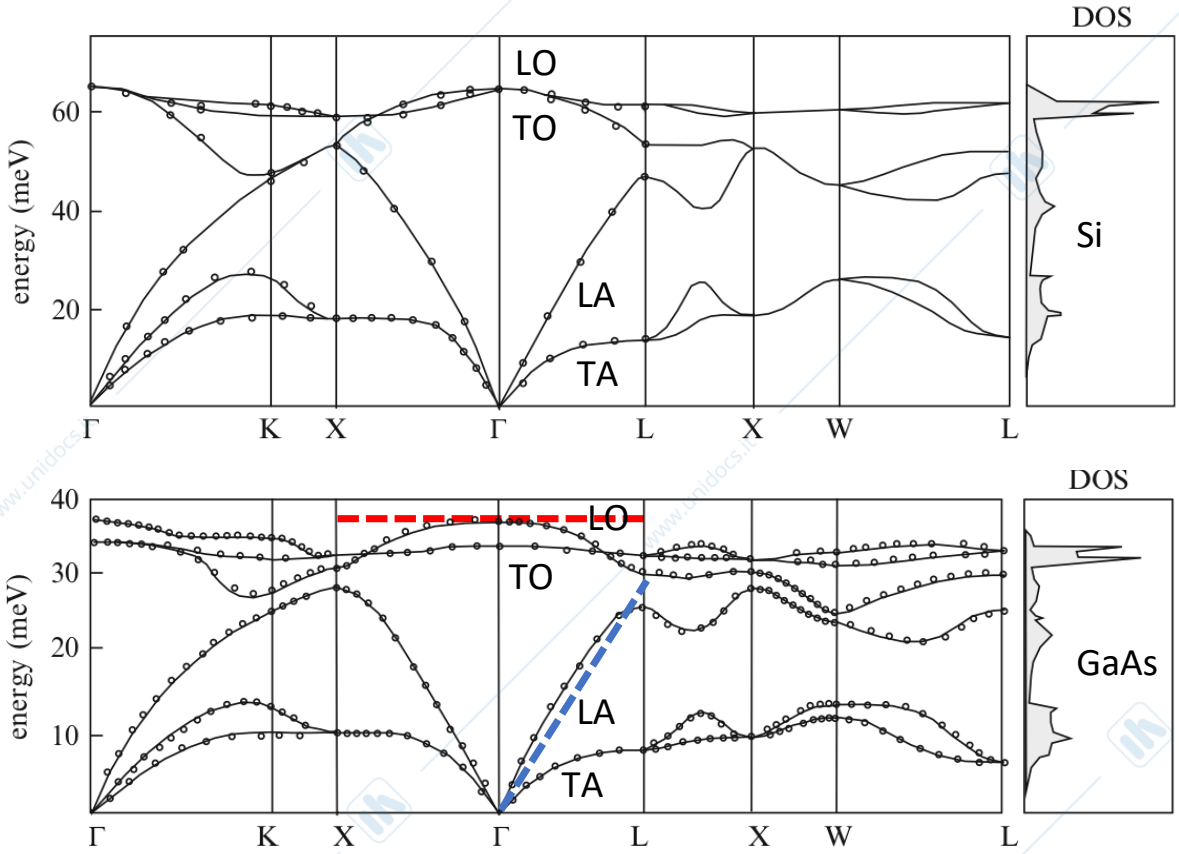


Figure 2.3.: Phonon dispersion curves and density of states in Si and GaAs calculated by Giannozzi et al [2, Ch. 5] (lines). The points are experimental results. Letters indicate the points of high symmetry in the Brillouin zone. Longitudinal and transeverse acoustic (optical) modes are denoted by LA, TA (LO, TO). Si and GaAs have two atoms per unit cell and therefore have three acoustic branches (two transverse and one longitudinal) and three optical branches. Transerve modes are degnerate along the high symmetry directions ΓL, ΓX. Non polar crystals, such as Si, can be described as consisting of netral atoms interacting with short range forces, which extend only to nearest neighbors, or just few nearest neighbors, and become negligible further away. In polar crystals, the crystal lattice can be described in terms of charged ions interacting with both short range forces and long-range Coulomb forces. The inclusion of long-range interactions in the interatomic forces causes a singularity (non-analyticity) in the dynamical matrix, which is responsible of the transverse-longitudinal splitting of the optical modes at $q \approx 0$. This splitting is the signature of the long-range nature of the interatomic forces. The density of states of the phonons (right panels) is defined in exactly the same way as the electronic densities of states, i.e., as the number of phonon modes per unit frequency per unit volume of real space. The phonon DOS shows peaks where the dispersion relation is flat. In Monte Carlo carrier transport simulations, the LO branch contributing to polar optical scattering is usually approximated as dispersionless (dashed red line), and the LA branch contributing to deformation potential scattering is approximated with a linear curve (dashed blue line).

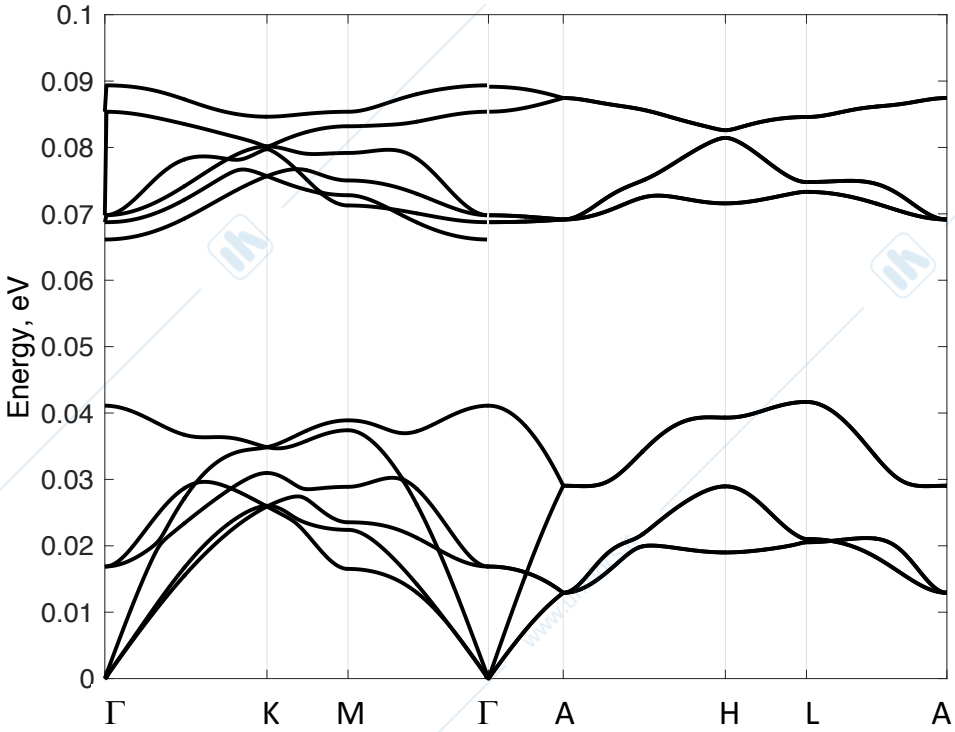


Figure 2.4.: Calculated phonon dispersion curves for wurtzite GaN obtained using the linear-response method within density-functional perturbation theory. Letters indicate the points of high symmetry in the Brillouin zone. Wurzite semiconductors with 4 atoms per unit cell support 12 mixed longitudinal and transverse phonon modes (nine optical and three acoustic modes) [15]. The ground state properties were obtained by self-consistent numerical minimization of the total energy with respect to the unit cell structural parameters, thus obtaining the dynamical matrix whose diagonalization leads to the phonon spectra.

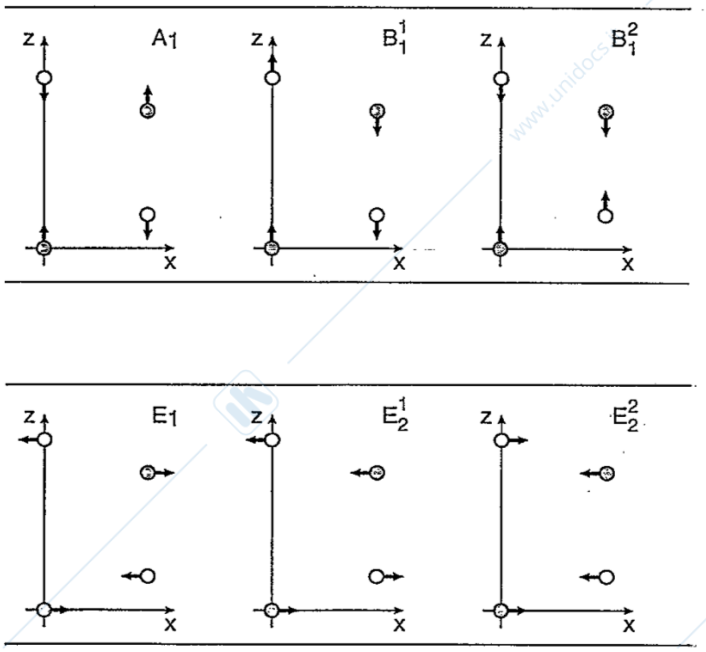


Figure 2.5.: Zone-center optical phonon modes in wurtzite crystals (acoustic modes are just translations in the long-wavelength limit). Reproduced from [16].

oscillating electric dipoles are called *optically active* as they may couple with electromagnetic fields (with appropriate frequencies). [5]

The dynamical matrix is the basic quantity for the calculations of the dispersion relations of the vibrational modes of any crystal. It yields the potential energy of the crystal for any configuration of the atomic displacements, and its eigenvalues are strictly related to the frequency of each mode. Force constants can be computed from first principles by means of density functional theory, or by empirical models (e.g., the shell model and the valence force field method [17]). As an example, the lattice dynamics of Si, GaAs, and wurtzite GaN computed by linear response density functional theory are reported in Figs. 2.3, 2.4.

2.0.4. Quantum theory of lattice vibrations

Until now, we have discussed lattice vibrations by means of classical mechanics. Here, we reconsider the problem from a quantum mechanical point of view, by replacing the familiar conjugate variables of position and momentum by creation and annihilation operators. Were we just concerned with the lattice dynamics, we could stop here since the quantum and classical formulations are completely equivalent as far as dispersion curves are concerned. As we will see soon, the motivation to move to a quantum mechanical treatment lies in the analysis of carrier-phonon interactions: in scattering theory, the creation operator corresponds to the emission of a phonon by the scattered electron; correspondingly, the absorption of a phonon by the electron corresponds to the annihilation of a phonon.

The translational invariance of the system allows us to write a special solution of the equation of motion (2.51) for the displacement of ion $l\gamma$ as (Bloch theorem)

$$\underline{u}_{l\gamma}^{\underline{q}\eta}(t) = \frac{1}{\sqrt{N_c M_\gamma}} \underline{\xi}_{\underline{q}\gamma}^{(\eta)} e^{i\underline{q}\cdot\underline{R}_{l\gamma}} e^{-i\omega_{\underline{q}\eta} t}, \tag{2.58}$$

where N_c is the number of cells in the crystal volume, and η is the phonon branch. The general solution can be expressed by a superposition of Bloch waves where the summation extends over all modes $\underline{q}\eta$

$$\underline{u}_{l\gamma}(t) = \frac{1}{\sqrt{N_c M_\gamma}} \sum_{\underline{q}\eta} \underline{\xi}_{\underline{q}\gamma}^{(\eta)} e^{i\underline{q}\cdot\underline{R}_{l\gamma}} e^{-i\omega_{\underline{q}\eta} t}. \tag{2.59}$$

More formally, normalizing the polarization vectors $\underline{\xi}_{\gamma\underline{q}}$ to obtain the unit vectors $\underline{e}_{\gamma\underline{q}}^{(\eta)}$, the general solution can be written in terms of complex *normal coordinates* [6] $U_{\underline{q}\eta}(t) = U_{\underline{q}\eta}(0) e^{-i\omega_{\underline{q}\eta} t}$

$$\underline{u}_{l\gamma}(t) = \frac{1}{\sqrt{N_c M_\gamma}} \sum_{\underline{q}\eta} U_{\underline{q}\eta}(t) e^{i\underline{q}\cdot\underline{R}_{l\gamma}} \underline{e}_{\underline{q}\gamma}^{(\eta)}. \tag{2.60}$$

We can now define a conjugate momentum $P_{\underline{q}\eta} = \dot{U}_{\underline{q}\eta}$

$$\underline{p}_{l\gamma}(t) = M_\gamma \dot{\underline{u}}_{l\gamma}(t) = \sqrt{\frac{M_\gamma}{N_c}} \sum_{\underline{q}\eta} P_{\underline{q}\eta}(t) e^{i\underline{q}\cdot\underline{R}_{l\gamma}} \underline{e}_{\underline{q}\gamma}^{(\eta)}. \tag{2.61}$$

For real displacement fields ($\underline{u}_{\underline{q}\gamma} = \underline{u}_{\underline{q}\gamma}^*$), we have

$$\underline{u}_{l\gamma}(t) = \frac{1}{\sqrt{N_c M_\gamma}} \sum_{\underline{q}\eta} U_{\underline{q}\eta}(t) e^{i\underline{q}\cdot\underline{R}_{l\gamma}} \underline{e}_{\underline{q}\gamma}^{(\eta)} = \frac{1}{\sqrt{N_c M_\gamma}} \sum_{\underline{q}\eta} U_{\underline{q}\eta}^*(t) e^{-i\underline{q}\cdot\underline{R}_{l\gamma}} \underline{e}_{\underline{q}\gamma}^{(\eta)*} = \underline{u}_{\underline{q}\gamma}^*. \tag{2.62}$$

[5] In polar crystals, the coupling of optical vibrational branches with the electromagnetic field leads to the concept of mixed phonon-photon quasiparticles, known as *polaritons*. Photons and transverse phonons strongly interact in a small range of the Brillouin zone from $q = 0$ to approximately the wavenumber that corresponds to the crossing of the (uncoupled) dispersion curves ($\omega_q = cq/\sqrt{\varepsilon_\infty}$ for photons), and determine the polariton dispersion curves.

[6] One may express the normal coordinates $U_{\underline{q}\eta}$ as a function of the atomic displacement coordinates $\underline{u}_{l\gamma}$ by inverting (2.60), which shows that the normal coordinates are actually *collective* coordinates, because they are superpositions of all local atomic displacement coordinates $\underline{u}_{l\gamma}$.

In the second sum, $\underline{q}$ can be replaced with $-\underline{q}$, and the sum runs over the same values because for each $\underline{q}$ there is a $-\underline{q}$, so

$$\underline{u}_{l\gamma}(t) = \frac{1}{\sqrt{N_c M_\gamma}} \sum_{\underline{q}\eta} U_{\underline{q}\eta}(t)\, e^{i\underline{q}\cdot\underline{R}_{l\gamma}}\, \underline{e}^{(\eta)}_{\underline{q}\gamma} = \frac{1}{\sqrt{N_c M_\gamma}} \sum_{\underline{q}\eta} U^*_{-\underline{q}\eta}(t)\, e^{i\underline{q}\cdot\underline{R}_{l\gamma}}\, \underline{e}^{(\eta)*}_{-\underline{q}\gamma} = \underline{u}^*_{\underline{q}\gamma}. \tag{2.63}$$

Translational invariance implies $\underline{e}^{(\eta)*}_{-\underline{q}\gamma} = \underline{e}^{(\eta)}_{\underline{q}\gamma}$ [14, Chapter 2], so

$$U^*_{\underline{q}\eta} = U_{-\underline{q}\eta}, \quad P^*_{\underline{q}\eta} = P_{-\underline{q}\eta}. \tag{2.64}$$

Then, the ion Hamiltonian

$$H^{\text{eff}}_{\text{ions}} = \frac{1}{2}\left(\sum_{l\gamma} \frac{|\underline{p}_{l\gamma}|^2}{M_\gamma} + \sum_{l\gamma l'\gamma'} \underline{u}_{l\gamma} \cdot \underline{\underline{D}}_{l\gamma l'\gamma'} \cdot \underline{u}_{l'\gamma'} \right) \tag{2.65}$$

can be written in terms of normal coordinates [14, Appendix I]

$$H^{\text{eff}}_{\text{ions}} = \frac{1}{2} \sum_{\underline{q}\eta} \left(\left| P_{\underline{q}\eta} \right|^2 + \omega^2_{\underline{q}\eta} \left| U_{\underline{q}\eta} \right|^2 \right) \tag{2.66}$$

where we have used the lattice sum rule [14, Appendix B], [3, Appendix A.2]

$$\sum_l e^{i(\underline{q}-\underline{q}')\cdot\underline{R}_l} = N_c \delta_{\underline{q},\underline{q}'} \tag{2.67}$$

and the orthonormality of the polarization vectors

$$\sum_\gamma \underline{e}^{(\eta)}_{\gamma\underline{q}} \cdot \underline{e}^{(\eta')*}_{\gamma\underline{q}} = \delta_{\eta\eta'}. \tag{2.68}$$

The main achievement of introducing normal coordinates is that the ionic Hamiltonian is now the sum of independent harmonic oscillators with frequency $\omega_{\underline{q}\eta}$, a form much simpler than (2.65). We now define the variable $a_{\underline{q}\eta}$ as a linear combination of $U_{\underline{q}\eta}$ and $P_{\underline{q}\eta}$

$$a_{\underline{q}\eta} = \frac{1}{\sqrt{2\hbar\omega_{\underline{q}\eta}}} \left(\omega_{\underline{q}\eta} U_{\underline{q}\eta} + i P_{\underline{q}\eta} \right) \tag{2.69a}$$

$$a^*_{\underline{q}\eta} = \frac{1}{\sqrt{2\hbar\omega_{\underline{q}\eta}}} \left(\omega_{\underline{q}\eta} U^*_{\underline{q}\eta} - i P^*_{\underline{q}\eta} \right). \tag{2.69b}$$

The normal coordinates may be expressed in terms of these new variables as

$$U_{\underline{q}\eta} = \sqrt{\frac{\hbar}{2\omega_{\underline{q}\eta}}} (a_{\underline{q}\eta} + a^*_{-\underline{q}\eta}) \tag{2.70a}$$

$$P_{\underline{q}\eta} = -i\sqrt{\frac{\hbar\omega_{\underline{q}\eta}}{2}} (a_{\underline{q}\eta} - a^*_{-\underline{q}\eta}). \tag{2.70b}$$

Inserting these expressions in (2.66) gives

$$\begin{aligned} H^{\text{eff}}_{\text{ions}} &= \frac{1}{2} \sum_{\underline{q}\eta} \left[\frac{\hbar\omega_{\underline{q}\eta}}{2} (a_{\underline{q}\eta} - a^*_{-\underline{q}\eta})(a^*_{\underline{q}\eta} - a_{-\underline{q}\eta}) + \omega^2_{\underline{q}\eta} \frac{\hbar}{2\omega_{\underline{q}\eta}} (a_{\underline{q}\eta} + a^*_{-\underline{q}\eta})(a^*_{\underline{q}\eta} + a_{-\underline{q}\eta}) \right] \\ &= \frac{1}{2} \sum_{\underline{q}\eta} \hbar\omega_{\underline{q}\eta} \left[a_{\underline{q}\eta} a^*_{\underline{q}\eta} + a^*_{\underline{q}\eta} a_{\underline{q}\eta} \right], \end{aligned} \tag{2.71}$$

which is already in the desired symmetrized form. At this point, we move to the realm of quantum mechanics by promoting the variables $a_{\underline{q}\eta}$ and $a^*_{\underline{q}\eta}$ to the corresponding annihilation and creation operators $\hat{a}_{\underline{q}\eta}$, $\hat{a}^\dagger_{\underline{q}\eta}$. Since the transformation is canonical, the commutation rules are preserved

$$[\hat{a}_{\underline{q}\eta}, \hat{a}^\dagger_{\underline{q}'\eta'}] = \delta_{\underline{q},\underline{q}'}\delta_{\eta,\eta'} \tag{2.72a}$$

$$[\hat{a}_{\underline{q}\eta}, \hat{a}_{\underline{q}'\eta'}] = [\hat{a}^\dagger_{\underline{q}\eta}, \hat{a}^\dagger_{\underline{q}'\eta'}] = 0. \tag{2.72b}$$

Insertion of the commutation relation (2.72a) in (2.71) gives the ionic Hamiltonian as a collection of independent harmonic oscillators

$$\hat{H}^{\text{eff}}_{\text{ions}} = \sum_{\underline{q}\eta} \hbar\omega_{\underline{q}\eta} \left(\hat{a}^\dagger_{\underline{q}\eta}\hat{a}_{\underline{q}\eta} + \frac{1}{2} \right). \tag{2.73}$$

For further reference, we obtain two equivalent forms of the displacement $\hat{\underline{u}}_{l\gamma}(\underline{r})$ replacing (2.70a) in (2.60)

$$\hat{\underline{u}}_{l\gamma} = \sum_{\underline{q}\eta} \sqrt{\frac{\hbar}{2N_c M_\gamma \omega_{\underline{q}\eta}}} \left(\hat{a}_{\underline{q}\eta} + \hat{a}^\dagger_{-\underline{q}\eta} \right) e^{i\underline{q}\cdot\underline{R}_{l\gamma}} \underline{e}^{(\eta)}_{\underline{q}\gamma} \tag{2.74a}$$

$$= \sum_{\underline{q}\eta} \sqrt{\frac{\hbar}{2N_c M_\gamma \omega_{\underline{q}\eta}}} \left(\hat{a}_{\underline{q}\eta} e^{i\underline{q}\cdot\underline{R}_{l\gamma}} + \hat{a}^\dagger_{\underline{q}\eta} e^{-i\underline{q}\cdot\underline{R}_{l\gamma}} \right) \underline{e}^{(\eta)}_{\underline{q}\gamma}. \tag{2.74b}$$

By definition, $\hat{a}_{\underline{q}}$ and $\hat{a}^\dagger_{\underline{q}}$ are not Hermitian, i.e., they do not correspond to observables, but the number operator $\hat{n}_{\underline{q}} = \hat{a}^\dagger_{\underline{q}}\hat{a}_{\underline{q}}$ is Hermitian: it can be shown that the eigenvalues $n_{\underline{q}}$ of $\hat{n}_{\underline{q}}$ are positive integers for bosons [2, Appendix C]

$$\hat{n}_{\underline{q}}|n_{\underline{q}}\rangle = n_{\underline{q}}|n_{\underline{q}}\rangle, \quad n_{\underline{q}} = 0, 1, 2, \ldots \tag{2.75}$$

The energy of the quantum state $|n_{\underline{q}}\rangle$ is given by $n_{\underline{q}}$ multiples of $\hbar\omega_{\underline{q}}$

$$\hat{H}^{\text{eff}}_{\text{ions}}|n_{\underline{q}}\rangle = \sum_{\underline{q}} \hbar\omega_{\underline{q}} \left(n_{\underline{q}} + \frac{1}{2} \right) |n_{\underline{q}}\rangle, \tag{2.76}$$

which means that there are $n_{\underline{q}}$ phonons in the state. The following properties can be obtained directly from the commutation relations (2.72) [11]

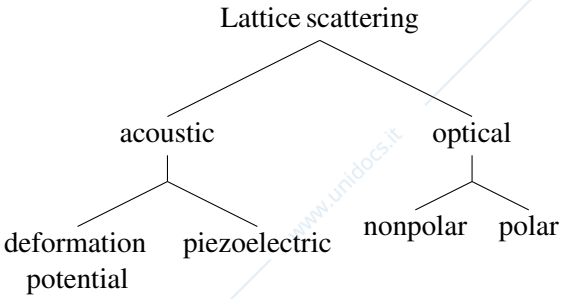
$$\hat{a}^\dagger_{\underline{q}}|n_{\underline{q}}\rangle = \sqrt{n_{\underline{q}}+1}\,|n_{\underline{q}}+1\rangle \tag{2.77a}$$

$$\hat{a}_{\underline{q}}|n_{\underline{q}}+1\rangle = \sqrt{n_{\underline{q}}+1}\,|n_{\underline{q}}\rangle. \tag{2.77b}$$

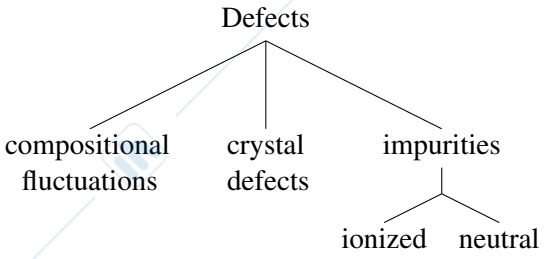
The previous analysis suggests to describe the vibrational properties of a semiconductor bulk crystal in terms of elementary excitations or quasiparticles called phonons, characterized by a three-dimensional wavevector $\underline{q}$, a corresponding energy dispersion $\hbar\omega_{\underline{q}\eta}$, and branch index η.

3. Scattering theory

Transitions between Bloch states (1.5) may occur due to several scattering mechanisms. If a full-band (analytical) description of the band structure is adopted, they may be classified as intraband (intravalley) or interband (intervalley) depending if initial and final states lie in the same bands (valleys). A more detailed classification is reported here depending on the interaction process.

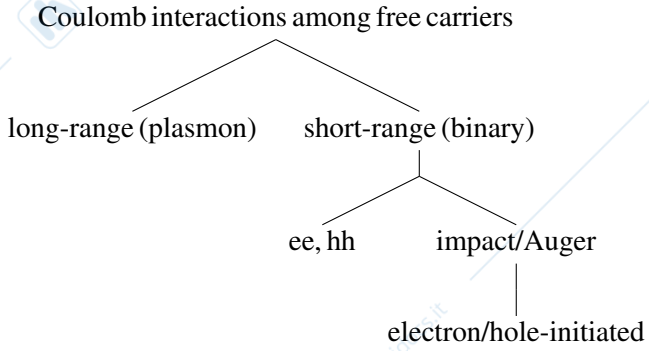
1. 

It consists in the absorption or emission of a phonon, which may be either acoustic or optical. For both phonon branches, there are two types of interactions: *electrostatic* in ionic (polar) semiconductors, or assisted by *deformation potential* active in all crystals. The latter is due to the variation of the band edge produced by the deformation of the lattice and is called deformation potential scattering when it is assisted by acoustic phonons, nonpolar optical scattering when it is assisted by optical phonons. The electrostatic interaction is due to the partially ionic nature of the bonding in zincblende crystals. For example, in GaAs, the reference material for this laboratory, the As atoms contribute more electrons to the bond than the Ga atoms. In other words, the electrons in the covalent bonds spend, on average, somewhat more time near the As atoms than the Ga atoms, so the As atoms are slightly negatively charged, while the Ga atoms are slightly positively charged. When the atoms move against each other, they create microscopic dipole fields that add up to a macroscopic electric field. This electric field can in turn scatter carriers. The electrostatic interaction may involve optical phonons (polar optical scattering) or acoustic phonons (piezoelectric scattering). From the point of view of the energy exchange, we shall see that when an optical phonon is involved, the scattering process is *inelastic*, while acoustic phonons carry a very small amount of energy, which allows to consider the process *elastic* at room temperature, see Fig. 2.3.

2. 

Defects may arise owing to dislocations of different types and geometries, to impurities, or to the random distribution of atoms among available lattice sites in alloys. The impurities can be ionized impurities, interacting with electrons through a long-range Coulomb field, or neutral impurities with a short-range interaction. Owing to the large mass of the impurities, this kind of scattering is always considered elastic. The effect of the impurities becomes more important at lower temperatures, when phonons become less effective. Impurities are always present in semiconductors, even when not introduced on purpose as dopants. Within the virtual crystal approximation (VCA), the

alloy is described as a uniform crystal, made of virtual atoms with (pseudo)potentials given by a concentration-weighted average of the (pseudo)potentials of the real atoms. What is not included in the VCA is the fluctuation of the crystal potential due to the random positions of the atoms. The microscopic variations with respect to the average crystal potential defined by the VCA produces a perturbation (alloy scattering) that generates transitions between the Bloch states of the virtual crystal.

3. 

Electron-electron collisions originate from fluctuations of the (Hartree or Hartree-Fock) mean field. Depending on the magnitude of the momentum transfer q, two different scenarios are possible: for small values of q, the interaction may involve collective excitations of carriers known as plasmons (long-range Coulomb interaction); for large values of q, plasmons decay into single-particle excitations, and the scattering process becomes a binary process between two electrons (short-range Coulomb interaction). Impact ionization belongs to the latter category and it can be classified as an interband process, since the energies exchange between the two interacting particles exceeds the energy of the band gap, so that an electron can be excited from the valence band to the conduction band, creating an electron-hole pair. The multiplication of carriers by impact ionization is of central importance in the theory of semiconductor devices both as a limiting mechanism and as a basis of device functionality.

3.1. Lattice scattering

A Bloch electron, conceived as an eigenstate propagating in the periodic potential defined by the periodic arrangement of the ions constituting the crystal, may be scattered to another Bloch state by lattice vibrations, which displace the position of neighboring atoms in the same direction (acoustic phonons) or in opposite directions (optical phonons). The interaction of phonons with the carriers may be due either to the deformation of the otherwise perfect crystal (deformation potential scattering) or to the electrostatic forces produced by the polarization waves that accompany the phonons (polar optical scattering). In the following we will be considering scattering by bulk phonons. [1]

To study the transitions of an electron between different Bloch states in a crystal, due to a perturbation, one starts with the assumption that the system can be separated into the electron of interest and the rest of the crystal. The Hamiltonian may be written as

$$\hat{H} = \underbrace{\hat{H}_{\mathrm{e}}^{\mathrm{eff}} + \hat{H}_{\mathrm{ions}}}_{\hat{\mathrm{H}}_0} + \hat{H}' \tag{3.1}$$

where $\hat{H}_{\mathrm{e}}^{\mathrm{eff}}$ represents the Hamiltonian of the electron in the perfect crystal (with the ions frozen in their equilibrium positions $\underline{R}_{j0}$), $\hat{H}_{\mathrm{ions}}$ is the Hamiltonian of the rest of the crystal, and $\hat{H}'$ is the perturbation Hamiltonian (2.6) that describes the interaction between the two subsystems. The Hamiltonian $\hat{H}_0$ of the unperturbed system is given by the first two terms in (3.1). Its eigenstates can be written as the direct product (2.7)

$$|\underline{k}n, \{n_{\underline{q}}\}\rangle = |\underline{k}n\rangle |\{n_{\underline{q}}\}\rangle, \tag{3.2}$$

where $|\underline{k}n\rangle$ and $|\{n_{\underline{q}}\}\rangle$ represent the unperturbed states of the electron and of the crystal, respectively. $|\{n_{\underline{q}}\}\rangle = |n_{\underline{q}_1}, n_{\underline{q}_2}, \cdots\rangle$ is a compact notation to indicate a collection of phonons. The transition probability per unit time from a state $|\underline{k}n, \{n_{\underline{q}}\}\rangle$ to a state $|\underline{k}'n', \{n'_{\underline{q}}\}\rangle$ induced by the perturbation Hamiltonian $\hat{H}'$ is given, to the lowest order, by Fermi's Golden rule (see Appendix B)

$$S(\underline{k}n, \{n_{\underline{q}}\}; \underline{k}'n', \{n'_{\underline{q}}\}) = \frac{2\pi}{\hbar} |\langle \underline{k}'n', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}n, \{n_{\underline{q}}\}\rangle|^2 \delta[E(\underline{k}'n', \{n'_{\underline{q}}\}) - E(\underline{k}n, \{n_{\underline{q}}\})]. \tag{3.3}$$

In general, $\hat{H}'$ can be written in Fourier series

$$\hat{H}' = \sum_{\underline{q}} \left[\hat{A}(\underline{q}) e^{\mathrm{i}\underline{q}\underline{r}} + \hat{A}^{\dagger}(\underline{q}) e^{-\mathrm{i}\underline{q}\underline{r}}\right], \tag{3.4}$$

where the sum of the two terms is needed to ensure the Hermeticity of $\hat{H}'$. In the particular case of acoustic deformation potential (Section 3.1.1) and polar optical scattering (Section 3.1.3)

$$\boxed{\hat{H}' = \mathrm{i} \sum_{\underline{q}} g_{\underline{q}} \left(\hat{a}_{\underline{q}} e^{\mathrm{i}\underline{q}\cdot\underline{r}} - \hat{a}^{\dagger}_{\underline{q}} e^{-\mathrm{i}\underline{q}\cdot\underline{r}}\right)} \tag{3.5}$$

with

$$\boxed{g_{\underline{q}} = \sum_{ij} \Xi_{ij} \sqrt{\frac{\hbar}{2\Omega\rho\omega_{\underline{q}}}} q_j [\underline{e}_{\underline{q}}]_i} \tag{3.6}$$

[1] Phonon waves in a nanotructure can be reflected back and forth by mechanical barriers, setting up standing phonon waves, thus leading to phonon confinement, exactly in the same way as electrons are confined by potential barriers. Just as electron confinement quantizes the electron's energy spectrum and therefore alters the electronic structure, phonon confinement can alter the dispersion relation. (This will have an effect on the electron-phonon scattering rates since they usually depend on the joint electron-phonon density of states and the phonon density of states will be altered if the phonon dispersion relation changes.) For example, in a superlattice consisting of two alternating semiconductors, acoustic phonon branches backfold in the Brillouin zone, exactly as discussed in Fig. 2.1, while optical branches of the two materials that do not overlap energetically may become localized in the respective layers. In quaternary systems of alternating layers, new modes, which do not exist in neither of the parent compounds, may appear at the interface, with amplitudes decaying away from the interface.

for the deformation potential interaction, and

$$\boxed{g_{\underline{q}} = \sqrt{\frac{e^2\hbar\omega_{LO}}{2\Omega}\left(\frac{q}{q^2+q_0^2}\right)^2\left(\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_s}\right)}}. \tag{3.7}$$

for the electrostatic interaction. Keeping only the first term in (3.4), the matrix element becomes

$$\langle \underline{k}'n', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}n, \{n_{\underline{q}}\}\rangle = \sum_{\underline{q}} \langle \{n'_{\underline{q}}\}|\hat{A}(\underline{q})|\{n_{\underline{q}}\}\rangle \int \mathrm{d}\underline{r}\left[\psi^{(n')}_{\underline{k}'}(\underline{r})\right]^* e^{\mathrm{i}\underline{q}\underline{r}}\psi^{(n)}_{\underline{k}}(\underline{r}) \tag{3.8}$$

where $\psi_{\underline{k}}(\underline{r})$ are the Bloch wavefunctions (the eigenfunctions of $\hat{H}^{\mathrm{eff}}_{\mathrm{e}}$). Let us split the integral above as a sum of integrals over the crystal cells labeled by the direct-lattice vectors $\underline{R}_l$. Then, remembering the form (1.5) of the Bloch functions

$$\psi^{(n)}_{\underline{k}}(\underline{r}) = \frac{1}{\sqrt{\Omega}}e^{\mathrm{i}\underline{k}\cdot\underline{r}}u^{(n)}_{\underline{k}}(\underline{r}) = \frac{1}{\sqrt{\Omega}}e^{\mathrm{i}\underline{k}\cdot\underline{r}}\sum_{\underline{G}} u^{(n)}_{\underline{G},\underline{k}}e^{\mathrm{i}\underline{G}\cdot\underline{r}} \tag{3.9}$$

we have

$$\langle \underline{k}'n', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}n, \{n_{\underline{q}}\}\rangle = \frac{1}{\Omega}\sum_{\underline{q}}\langle \{n'_{\underline{q}}\}|\hat{A}(\underline{q})|\{n'_{\underline{q}}\}\rangle \int_\Omega \mathrm{d}\underline{r}\sum_{\underline{G}\underline{G}'} u^{(n')*}_{\underline{G}'\underline{k}'}u^{(n)}_{\underline{G}\underline{k}}e^{\mathrm{i}(\underline{k}-\underline{k}'+\underline{G}-\underline{G}'+\underline{q})\cdot\underline{r}}. \tag{3.10}$$

Replacing $\underline{r}$ with the sum of a lattice vector $\underline{R}_l$ and a coordinate $\underline{\rho}$ spanning a single cell Ω_{c} (where N_{c} is the number of the unit cells in the crystal, and $\Omega = N_{\mathrm{c}}\Omega_{\mathrm{c}}$)

$$\underline{r} \to \underline{\rho} + \underline{R}_l \tag{3.11}$$

$$\int_\Omega \mathrm{d}\underline{r} \to \sum_l \int_{\Omega_{\mathrm{c}}} \mathrm{d}\underline{\rho} \tag{3.12}$$

we obtain

$$\langle \underline{k}'n', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}n, \{n_{\underline{q}}\}\rangle =$$
$$\frac{1}{\Omega}\sum_{\underline{q}}\langle \{n'_{\underline{q}}\}|\hat{A}(\underline{q})|\{n_{\underline{q}}\}\rangle \sum_l e^{\mathrm{i}(\underline{k}-\underline{k}'+\underline{q}+\underline{G}-\underline{G}')\cdot\underline{R}_l}\int_{\Omega_{\mathrm{c}}}\mathrm{d}\underline{\rho}\sum_{\underline{G}\underline{G}'}u^{(n')*}_{\underline{G}'\underline{k}'}u^{(n)}_{\underline{G}\underline{k}}e^{\mathrm{i}(\underline{k}-\underline{k}'+\underline{G}-\underline{G}'+\underline{q})\cdot\underline{\rho}}. \tag{3.13}$$

The sum over lattice sites is

$$\sum_l e^{\mathrm{i}(\underline{k}-\underline{k}'+\underline{q}+\underline{G}-\underline{G}')\cdot\underline{R}_l} = N_{\mathrm{c}}\sum_{\underline{G}''}\delta(\underline{k}-\underline{k}'-\underline{G}''+\underline{q}), \tag{3.14}$$

i.e., the argument of the complex exponential must be some reciprocal-lattice vector $\underline{G}'' = \underline{G}-\underline{G}'$ in order to obtain a non-vanishing contribution. The Kronecker function in this equation describes the conservation of crystal momentum in electronic scattering in crystals, verified up to a reciprocal lattice vector. Thus, with some renaming of the dummy $\underline{G}$-vectors, we can write

$$\langle \underline{k}'n', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}n, \{n_{\underline{q}}\}\rangle = \sum_{\underline{q}}\langle \{n'_{\underline{q}}\}|\hat{A}(\underline{q})|\{n_{\underline{q}}\}\rangle\sum_{\underline{G}}\delta(\underline{k}-\underline{k}'+\underline{q}-\underline{G})\underbrace{\left[\frac{N_{\mathrm{c}}}{\Omega}\int_{\Omega_{\mathrm{c}}}\mathrm{d}\underline{\rho}\sum_{\underline{G}'\underline{G}''}u^{(n')*}_{\underline{G}''\underline{k}'}u^{(n)}_{\underline{G}'\underline{k}}e^{\mathrm{i}(\underline{G}+\underline{G}'-\underline{G}'')\cdot\underline{\rho}}\right]}_{\text{overlap factor}}. \tag{3.15}$$

The last term in the square bracket is the overlap factor

$$\mathscr{G}_{\underline{k},\underline{k}',n,n';\underline{G}} = \frac{1}{\Omega_{\mathrm{c}}}\int_{\Omega_{\mathrm{c}}}\mathrm{d}\underline{\rho}\sum_{\underline{G}'\underline{G}''}u^{(n')*}_{\underline{G}''\underline{k}'}u^{(n)}_{\underline{G}'\underline{k}}e^{\mathrm{i}(\underline{G}+\underline{G}'-\underline{G}'')\cdot\underline{\rho}} = \sum_{\underline{G}'}u^{(n')*}_{\underline{G}'-\underline{G},\underline{k}'}u^{(n)}_{\underline{G}'\underline{k}} \tag{3.16}$$

since

$$\int_{\Omega_c} \mathrm{d}\underline{\rho} e^{\mathrm{i}\underline{G}\cdot\underline{\rho}} = \Omega_c \delta_{\underline{G},0}. \tag{3.17}$$

The Kronecker function $\delta(\underline{k}-\underline{k}'+\underline{q}-\underline{G})$ eliminates the sum over $\underline{q}$ in (3.15), selecting $\underline{q} = \underline{k}-\underline{k}'+\underline{G}$. Collecting the above results, the matrix element is

$$\langle \underline{k}'n', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}n, \{n_{\underline{q}}\}\rangle = \sum_{\underline{G}} \langle \{n'_{\underline{q}}\}|\hat{A}(\underline{k}-\underline{k}'+\underline{G})|\{n_{\underline{q}}\}\rangle \mathscr{G}_{\underline{k},\underline{k}',n,n';\underline{G}}. \tag{3.18}$$

Application of Fermi's Golden rule gives the transition rate

$$S\left(\underline{k}n, \{n_{\underline{q}}\}; \underline{k}'n', \{n'_{\underline{q}}\}\right) = \frac{2\pi}{\hbar} \left| \sum_{\underline{G}} \langle \{n'_{\underline{q}}\}|\hat{A}(\underline{k}-\underline{k}'+\underline{G})|\{n_{\underline{q}}\}\rangle \, \mathscr{G}_{\underline{k},\underline{k}',n,n';\underline{G}} \right|^2 \delta[E(\underline{k}'n', \{n'_{\underline{q}}\}) - E(\underline{k}n, \{n_{\underline{q}}\})]. \tag{3.19}$$

The scattering processes in which $\underline{G}$ is different from zero are called *umklapp*, or U-processes; if $\underline{G} = 0$ the process is called *non-umklapp* or N-process (*Normal* process).[2] Notice that contributions to the sum over $\underline{G}$ decrease quickly as the magnitude of the $\underline{G}$-vectors increases. This is due to the fact that the exponential inside the overlap factor (3.16) oscillates fast as $\underline{G}$ grows beyond the spectral range of the Bloch functions. Moreover, the magnitude of the matrix element usually decreases quickly at large $\underline{q}$. In these circumstances it is acceptable to consider only the smallest $\underline{G}$ i.e., $\underline{G} = 0$, corresponding to N-processes

$$S\left(\underline{k}n, \{n_{\underline{q}}\}; \underline{k}'n', \{n'_{\underline{q}}\}\right) = \frac{2\pi}{\hbar} \left| \langle \{n'_{\underline{q}}\}|\hat{A}(\underline{k}-\underline{k}')|\{n_{\underline{q}}\}\rangle \right|^2 |\mathscr{G}_{\underline{k},\underline{k}',n,n'}|^2 \delta[E(\underline{k}'n', \{n'_{\underline{q}}\}) - E(\underline{k}n, \{n_{\underline{q}}\})], \tag{3.20}$$

with the understanding that $\underline{k}-\underline{k}'$ must be mapped in the first Brillouin zone. The transition rates given by the Fermi Golden rule in (3.20) contain the matrix elements of the interaction Hamiltonian between the final and initial states of the total system. In some cases, the symmetry properties of the electron wavefunctions and of the interaction Hamiltonian are such that the matrix elements are zero, and the corresponding transitions are said to be forbidden by a selection rule.[3]

To proceed further, we need explicit forms for $\hat{H}'$, which depend on the particular scattering mechanism considered. To simplify the discussion we assume a simple model of the electronic structure, see Fig. 3.2.

[2] Intravalley transitions are usually N-processes because the distance between $\underline{k}$ and $\underline{k}'$ is small compared to the dimensions of the Brillouin zone. Intervalley transitions may be U-processes, because of the large values associated to $\underline{k}$ and $\underline{k}'$.

[3] The effects of symmetry can be formally analyzed using algebra and more specifically "group theory" [18]. A nonlethal introduction to this topic may be found in [17, A Pedestrian's Guide to Group Theory].

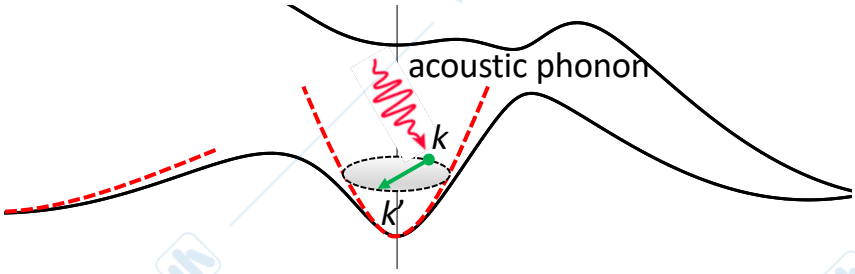
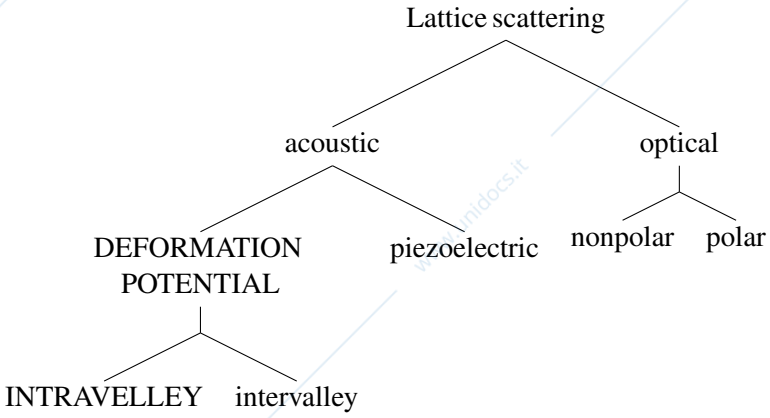


Figure 3.1.: Intravalley transition assisted by (elastic) acoustic scattering.



3.1.1. Deformation potential scattering by acoustic phonons (intravalley)

The deformation-potential interaction is due to the effect that an ionic displacement, associated with a phonon, has on the ionic potential and consequently on the electron energy. It is called a nonpolar interaction, because the electrons do not feel any long-range electric (dipole) field, but feel simply the perturbation caused by the fact that the ions in the electronic Hamiltonian are not frozen in their equilibrium positions. In principle both acoustic and optical phonons contribute to this interaction, but here we consider only deformation potential scattering by longitudinal ($\underline{e}_{\underline{q}} = \underline{q}/q$) acoustic waves (LA), dropping from now on the index η labeling the phonon branches. In the macroscopic limit ($N_c \to \infty$), we may replace ($l \to$ lattice site, $\gamma \to$ atom type)

$$\underline{R}_{l\gamma} \to \underline{r}, \tag{3.21}$$

with the understanding that $\underline{\hat{u}}(\underline{r})$ in (2.74b) becomes now the displacement of the unit cell

$$\underline{\hat{u}}_{l\gamma} = \sum_{\underline{q}\eta} \sqrt{\frac{\hbar}{2N_c M_\gamma \omega_{\underline{q}\eta}}} \left(\hat{a}_{\underline{q}\eta} e^{i\underline{q}\cdot\underline{R}_{l\gamma}} + \hat{a}^\dagger_{\underline{q}\eta} e^{-i\underline{q}\cdot\underline{R}_{l\gamma}} \right) \underline{e}^{(\eta)}_{\underline{q}\gamma} \tag{3.22}$$

$$\Downarrow$$

$$\underline{\hat{u}}(\underline{r}) = \sum_{\underline{q}} \sqrt{\frac{\hbar}{2\Omega\rho\,\omega_{\underline{q}}}} \left(\hat{a}_{\underline{q}} e^{i\underline{q}\cdot\underline{r}} + \hat{a}^\dagger_{\underline{q}} e^{-i\underline{q}\cdot\underline{r}} \right) \underline{e}_{\underline{q}}, \tag{3.23}$$

where $\rho = N_c M/\Omega$ is the density of the crystal, and $M = \sum_\gamma M_\gamma$ is the mass of the unit cell. The perturbation due to the motion of the ions, seen by an electron, may be written as a Taylor expansion about the equilibrium position of the ions. Assuming that the displacements of the ions from their equilibrium positions are small enough so that the use of linear response theory is justified, we may keep only the first-order term, ignoring higher-order terms that give rise to many-phonon processes

$$\hat{H}' = \sum_{ij} \Xi_{ij} \frac{\partial \hat{u}_i}{\partial r_j}, \tag{3.24}$$

where Ξ_{ij} is the deformation potential tensor ($i,j \to x,y,z$), and

$$\frac{\partial \hat{u}_i}{\partial r_j} = \mathrm{i}\sum_{\underline{q}} \sqrt{\frac{\hbar}{2\Omega\rho\,\omega_{\underline{q}}}} \left(\hat{a}_{\underline{q}} e^{\mathrm{i}\underline{q}\cdot\underline{r}} - \hat{a}^{\dagger}_{\underline{q}} e^{-\mathrm{i}\underline{q}\cdot\underline{r}}\right) q_j [\underline{e}_{\underline{q}}]_i. \tag{3.25}$$

By using the above expressions, we obtain the interaction Hamiltonian

$$\hat{H}' = \mathrm{i}\sum_{ij}\Xi_{ij}\sum_{\underline{q}} \sqrt{\frac{\hbar}{2\Omega\rho\,\omega_{\underline{q}}}} \left(\hat{a}_{\underline{q}} e^{\mathrm{i}\underline{q}\cdot\underline{r}} - \hat{a}^{\dagger}_{\underline{q}} e^{-\mathrm{i}\underline{q}\cdot\underline{r}}\right) q_j [\underline{e}_{\underline{q}}]_i, \tag{3.26}$$

which may be conveniently written as (3.5)

$$\hat{H}' = \mathrm{i}\sum_{\underline{q}} g_{\underline{q}} \left(\hat{a}_{\underline{q}} e^{\mathrm{i}\underline{q}\cdot\underline{r}} - \hat{a}^{\dagger}_{\underline{q}} e^{-\mathrm{i}\underline{q}\cdot\underline{r}}\right), \tag{3.27}$$

with

$$g_{\underline{q}} = \sum_{ij}\Xi_{ij}\sqrt{\frac{\hbar}{2\Omega\rho\,\omega_{\underline{q}}}} q_j [\underline{e}_{\underline{q}}]_i. \tag{3.28}$$

If the electron-phonon interaction is weak, we may assume that the state of the combined system is the direct product $|\underline{k},\{n_{\underline{q}}\}\rangle = |\underline{k}\rangle\,|\{n_{\underline{q}}\}\rangle$, where $|\underline{k}\rangle$ is a single-electron Bloch state (the band index is omitted as we are considering intra-valley processes) and $|\{n_{\underline{q}}\}\rangle$ describes a configurations of phonons, $n_{\underline{q}_i}$ of them in a single-phonon state $|\underline{q}_i\rangle$. Keeping in mind (3.8) and (3.18), the matrix element associated to the Hamiltonian (3.27) is

$$\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k},\{n_{\underline{q}}\}\rangle = \mathrm{i}\sum_{\underline{q}} g_{\underline{q}} \langle \{n'_{\underline{q}}\}|\hat{a}_{\underline{q}}|\{n_{\underline{q}}\}\rangle \langle \underline{k}'|e^{\mathrm{i}\underline{q}\cdot\underline{r}}|\underline{k}\rangle - \mathrm{i}\sum_{\underline{q}} g_{\underline{q}} \langle \{n'_{\underline{q}}\}|\hat{a}^{\dagger}_{\underline{q}}|\{n_{\underline{q}}\}\rangle \langle \underline{k}'|e^{-\mathrm{i}\underline{q}\cdot\underline{r}}|\underline{k}\rangle. \tag{3.29}$$

For a spherical parabolic valley, the overlap factor defined in (3.16) is equal to one, and we are left with (we consider only the normal term)

$$\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k},\{n_{\underline{q}}\}\rangle = \mathrm{i}\sum_{\underline{q}} g_{\underline{q}} \langle \{n'_{\underline{q}}\}|\hat{a}_{\underline{q}}|\{n_{\underline{q}}\}\rangle \delta_{\underline{k}-\underline{k}'+\underline{q}} - \mathrm{i}\sum_{\underline{q}} g_{\underline{q}} \langle \{n'_{\underline{q}}\}|\hat{a}^{\dagger}_{\underline{q}}|\{n_{\underline{q}}\}\rangle \delta_{\underline{k}-\underline{k}'-\underline{q}}. \tag{3.30}$$

Using the properties of the phonon creation and annihilation operators (2.77), it is easy to show that

$$\langle \{n'_{\underline{q}}\}|\hat{a}_{\underline{q}}|\{n_{\underline{q}}\}\rangle = \sqrt{n_{\underline{q}}}\;\delta_{n'_{\underline{q}}+1,n_{\underline{q}}} \tag{3.31a}$$

$$\langle \{n'_{\underline{q}}\}|\hat{a}^{\dagger}_{\underline{q}}|\{n_{\underline{q}}\}\rangle = \sqrt{n_{\underline{q}}+1}\;\delta_{n'_{\underline{q}},n_{\underline{q}}+1}, \tag{3.31b}$$

which means that the occupation of mode $\underline{q}$ can change by one unit only. Let's prove the first identity:

$$\begin{aligned} \hat{a}_{\underline{q}}|n_{\underline{q}}\rangle &= \sqrt{n_{\underline{q}}}\;|n_{\underline{q}}-1\rangle \\ \langle n'_{\underline{q}}|\hat{a}_{\underline{q}}|n_{\underline{q}}\rangle &= \sqrt{n_{\underline{q}}}\;\langle n'_{\underline{q}}|n_{\underline{q}}-1\rangle = \sqrt{n_{\underline{q}}}\;\delta_{n'_{\underline{q}},n_{\underline{q}}-1}. \end{aligned}$$

With the previous results (3.30) becomes

$$\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k},\{n_{\underline{q}}\}\rangle = \mathrm{i}\sum_{\underline{q}} g_{\underline{q}} \left[\sqrt{n_{\underline{q}}}\,\delta_{n'_{\underline{q}}+1,n_{\underline{q}}}\delta_{\underline{k}-\underline{k}'+\underline{q}} - \sqrt{n_{\underline{q}}+1}\,\delta_{n'_{\underline{q}},n_{\underline{q}}+1}\delta_{\underline{k}-\underline{k}'-\underline{q}}\right]. \tag{3.32}$$

So the squared matrix element

$$\begin{aligned} |\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k},\{n_{\underline{q}}\}\rangle|^2 &= \langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k},\{n_{\underline{q}}\}\rangle \left(\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k},\{n_{\underline{q}}\}\rangle\right)^* \\ &= \langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k},\{n_{\underline{q}}\}\rangle\;\langle \underline{k},\{n_{\underline{q}}\}|\hat{H}'|\underline{k}', \{n'_{\underline{q}}\}\rangle \end{aligned} \tag{3.33}$$

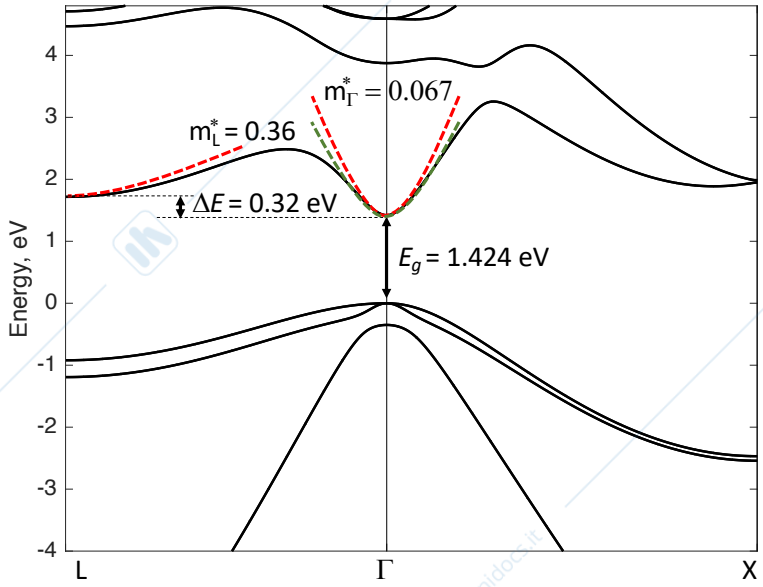


Figure 3.2.: Electronic structure of GaAs computed by the empirical pseudopotential method (solid lines), and its effective mass approximation near the Γ and L minima of the first conduction band (parabolic approximation, red dashed curves). The green curve is the Γ valley in the nonparabolic approximation.

reads

$$|\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}, \{n_{\underline{q}}\}\rangle|^2 = + \sum_{\underline{q}\underline{q}'} g_{\underline{q}} \sqrt{n_{\underline{q}}}\, \delta_{n'_{\underline{q}}+1, n_{\underline{q}}} \delta_{\underline{k}-\underline{k}'+\underline{q}}\, g_{\underline{q}'} \sqrt{n_{\underline{q}'}}\, \delta_{n'_{\underline{q}'}+1, n_{\underline{q}'}} \delta_{\underline{k}-\underline{k}'+\underline{q}'} \tag{3.34a}$$

$$+ \sum_{\underline{q}\underline{q}'} g_{\underline{q}} \sqrt{n_{\underline{q}}+1}\, \delta_{n'_{\underline{q}}, n_{\underline{q}}+1} \delta_{\underline{k}-\underline{k}'-\underline{q}}\, g_{\underline{q}'} \sqrt{n_{\underline{q}'}+1}\, \delta_{n'_{\underline{q}'}, n_{\underline{q}'}+1} \delta_{\underline{k}-\underline{k}'-\underline{q}'} \tag{3.34b}$$

$$- \sum_{\underline{q}\underline{q}'} g_{\underline{q}} \sqrt{n_{\underline{q}}}\, \delta_{n'_{\underline{q}}+1, n_{\underline{q}}} \delta_{\underline{k}-\underline{k}'+\underline{q}}\, g_{\underline{q}'} \sqrt{n_{\underline{q}'}+1}\, \delta_{n'_{\underline{q}'}, n_{\underline{q}'}+1} \delta_{\underline{k}-\underline{k}'-\underline{q}'} \tag{3.34c}$$

$$- \sum_{\underline{q}\underline{q}'} g_{\underline{q}} \sqrt{n_{\underline{q}}+1}\, \delta_{n'_{\underline{q}}, n_{\underline{q}}+1} \delta_{\underline{k}-\underline{k}'-\underline{q}}\, g_{\underline{q}'} \sqrt{n_{\underline{q}'}}\, \delta_{n'_{\underline{q}'}+1, n_{\underline{q}'}} \delta_{\underline{k}-\underline{k}'+\underline{q}'}. \tag{3.34d}$$

The last two terms are always zero ($\underline{q} = -\underline{q}'$). Consider for example the third term:

$$\delta_{\underline{k}-\underline{k}'+\underline{q}}\, \delta_{\underline{k}-\underline{k}'-\underline{q}'} \to \underline{q} = -\underline{q}'$$
$$\delta_{n'_{\underline{q}}+1, n_{\underline{q}}}\, \delta_{n'_{\underline{q}'}, n_{\underline{q}'}+1} = \delta_{n'_{\underline{q}}+1, n_{\underline{q}}}\, \delta_{n'_{-\underline{q}}, n_{-\underline{q}}+1} = 0,$$

since $\sum_{\underline{q}} = \sum_{-\underline{q}}$. A simular argument applies to the last term. So we are left with the first two terms that have $\underline{q} = \underline{q}'$

$$|\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}, \{n_{\underline{q}}\}\rangle|^2 = \sum_{\underline{q}} g^2_{\underline{q}} \left[\underbrace{n_{\underline{q}}\, \delta_{n'_{\underline{q}}+1, n_{\underline{q}}} \delta_{\underline{k}-\underline{k}'+\underline{q}}}_{\text{absorption}} + \underbrace{(n_{\underline{q}}+1)\, \delta_{n'_{\underline{q}}, n_{\underline{q}}+1} \delta_{\underline{k}-\underline{k}'-\underline{q}}}_{\text{emission}} \right]. \tag{3.35}$$

Applying the momentum conservation condition, we have

$$|\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}, \{n_{\underline{q}}\}\rangle|^2 = \underbrace{g^2_{\underline{q}_-}\, n_{\underline{q}_-}\, \delta_{n'_{\underline{q}_-}+1, n_{\underline{q}_-}}}_{\text{absorption}} + \underbrace{g^2_{\underline{q}_+}\, (n_{\underline{q}_+}+1)\, \delta_{n'_{\underline{q}_+}, n_{\underline{q}_+}+1}}_{\text{emission}}, \tag{3.36}$$

with where $\underline{q}_{\mp} = \mp(\underline{k} - \underline{k}')$ is wavevector of the phonon absorbed (upper sign) or emitted (lower sign) by the electron. (3.36) is usually written more compactly as

$$|\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}, \{n_{\underline{q}}\}\rangle|^2 = \sum_{\mp} g^2_{\underline{q}_{\mp}} \left(n_{\underline{q}_{\mp}} + \frac{1}{2} \mp \frac{1}{2} \right) \delta_{n'_{\underline{q}_{\mp}}, n_{\underline{q}_{\mp}} \mp 1}. \tag{3.37}$$

Substituting in Fermi's Golden rule

$$S_{\text{aco}}\left(\underline{k}, \{n_{\underline{q}}\}; \underline{k}', \{n'_{\underline{q}}\}\right) = \frac{2\pi}{\hbar} |\langle \underline{k}', \{n'_{\underline{q}}\}|\hat{H}'|\underline{k}, \{n_{\underline{q}}\}\rangle|^2 \delta\left(E(\underline{k}', \{n'_{\underline{q}}\}) - E(\underline{k}, \{n_{\underline{q}}\})\right). \tag{3.38}$$

Remembering that

$$E(\underline{k}, \{n_{\underline{q}}\}) = E(\underline{k}) + \sum_{\underline{q}} \hbar\omega_{\underline{q}} \left(n_{\underline{q}} + \frac{1}{2} \right), \tag{3.39}$$

we have

$$S_{\text{aco}}\left(\underline{k}, \{n_{\underline{q}}\}; \underline{k}', \{n'_{\underline{q}}\}\right) = \frac{2\pi}{\hbar} g^2_{\underline{q}_{\mp}} \left(n_{\underline{q}_{\mp}} + \frac{1}{2} \mp \frac{1}{2} \right) \delta_{n'_{\underline{q}_{\mp}}, n_{\underline{q}_{\mp}} \mp 1} \, \delta\left(E(\underline{k}') - E(\underline{k}) \mp \hbar\omega_{\underline{q}_{\mp}}\right). \tag{3.40}$$

The transition probability for the electron system only is obtained by taking the average over the state of the crystal

$$S_{\text{aco}}(\underline{k}, \underline{k}') = \langle S(\underline{k}, \{n_{\underline{q}}\}; \underline{k}', \{n'_{\underline{q}}\})\rangle_c = \frac{2\pi}{\hbar} g^2_{\underline{q}_{\mp}} \left(N_{\underline{q}} + \frac{1}{2} \mp \frac{1}{2} \right) \delta\left(E(\underline{k}') - E(\underline{k}) \mp \hbar\omega_{\underline{q}_{\mp}}\right), \tag{3.41}$$

where

$$N_{\underline{q}} = \langle n_{\underline{q}_{\mp}} \rangle_c = \frac{1}{e^{(\hbar\omega_{\underline{q}})/(k_B T)} - 1} \tag{3.42}$$

is the average number of phonons in mode $\underline{q}$ (or $-\underline{q}$, the symmetry of all crystal we consider makes this sign irrelevant) at thermal equilibrium, i.e., we assume that the crystal is not perturbed significantly by the few electrons moving around, which means that its temperature will not change too much (deviations from this condition lead to the problem of "hot phonons", see footnote 2). The physical interpretation of (3.41) is intuitively clear. The probability per unit time of absorbing a phonon is proportional to the number of phonons available to be absorbed, while the probability of emitting a phonon of wavevector $\underline{q}$ is proportional to $N_{\underline{q}} + 1$. The term $(+1)$ added to $N_{\underline{q}}$ in the emission case is due to the so-called *spontaneous emission*, a quantum phenomenon not present in classical physics, which describes the possibility of electron interaction with the phonon field (as well as with the electromagnetic field) even when phonons are not present. The term proportional to $N_{\underline{q}}$ gives the transition probability induced by the phonons already present in the crystal and is called *stimulated emission*.

The total rate $W(\underline{k})$ out of the Bloch state $\psi_{\underline{k}}$ can be evaluated numerically by integrating over the final states in the Brillouin zone, see the collision term defined in (1.3)

$$W_{\text{aco}}(\underline{k}) = \sum_{\underline{k}'} \frac{\pi}{\Omega\rho\,\omega_q} \underbrace{\left| \sum_{ij} \Xi_{ij} q_j [\underline{e}_{\underline{q}}]_i \right|^2}_{\mathrm{D_a^2 q^2}} \left(N_{\underline{q}} + \frac{1}{2} \mp \frac{1}{2} \right) \delta\left(E(\underline{k}') - E(\underline{k}) \mp \hbar\omega_{\underline{q}}\right) \tag{3.43}$$

The numerical evaluation of the scattering rates when the electron band structure is known numerically from, say, DFT or empirical pseudopotential calculations is a complex task. Interested readers may find numerical details in [4, Chapter 12]. For a parabolic band at the center of the Brillouin zone of a cubic semiconductor, the deformation-potential constant Ξ_{ij} in (3.28) is a second-rank diagonal tensor with equal diagonal elements and therefore can be treated as a scalar quantity. The squared factor that appears in (3.43) reduces to $D_a^2 q^2$ for longitudinal acoustic phonons ($\underline{e}_{\underline{q}} = \underline{q}/q$), while it vanishes for transverse

acoustic modes. If we approximate the longitudinal acoustic (LA) phonon dispersion as $\omega_q = q v_l$, where v_l is the longitudinal sound velocity, see (2.35) and Fig. 2.3, the integral over the Brillouin zone in (3.43) can be carried out analytically. Taking the continuum limit

$$\sum_{\underline{k}'} \to \frac{\Omega}{(2\pi)^3} \int \mathrm{d}\underline{k}', \tag{3.44}$$

and assuming the overlap integral equal to one or included in the phenomenological coupling constant D_a

$$W_{\mathrm{aco}}(\underline{k}) = \frac{\Omega}{(2\pi)^3} \int_{\mathrm{BZ}} \mathrm{d}\underline{k}' \frac{\pi D_a^2 q}{\Omega \rho v_l} \left(N_{\underline{q}} + \frac{1}{2} \mp \frac{1}{2} \right) \delta \left(E(\underline{k}') - E(\underline{k}) \mp \hbar \omega_{\underline{q}} \right). \tag{3.45}$$

Since the energy of the acoustic phonons involved in the transition is much smaller than the thermal energy at room temperature ($\hbar q v_l \ll k_B T$), the phonon population N_q is usually approximated by the equipartition expression (obtained with the Laurent expansion of the Bose distribution)

$$N_q = \frac{1}{e^{(\hbar \omega_{\underline{q}})/(k_B T)} - 1} \approx \frac{k_B T}{\hbar q v_l}. \tag{3.46}$$

For spherical parabolic bands, the conservation of energy reads (N-process)

$$\frac{\hbar^2 k'^2}{2m^*} - \frac{\hbar^2 k^2}{2m^*} \mp \hbar \omega_{\underline{q}} = 0, \tag{3.47}$$

where m^* is the effective mass. We may combine the two δ-functions expressing the conservation of momentum and energy into a single one

$$\begin{aligned} &\delta(\underline{k}' - \underline{k} \mp \underline{q})\, \delta(E(\underline{k}') - E(\underline{k}) \mp \hbar \omega_{\underline{q}}) = \\ &\delta(\underline{k}' - \underline{k} \mp \underline{q})\, \delta\left(\frac{\hbar^2 k'^2}{2m^*} - \frac{\hbar^2 k^2}{2m^*} \mp \hbar \omega_{\underline{q}}\right) = \qquad (3.48) \\ &\delta\left(\frac{\hbar^2 q^2}{2m^*} \pm \frac{\hbar^2 k q \cos\theta'}{m^*} \mp \hbar \omega_{\underline{q}}\right), \qquad (3.49) \end{aligned}$$

where θ' is the polar angle between $\underline{k}$ and $\underline{q}$, see Fig. 3.3. Neglecting the energy of the phonon (elastic approximation), we have ($N_{\underline{q}} \approx N_{\underline{q}} + 1$)

$$\begin{aligned} W_{\mathrm{aco}}(\underline{k}) &= \frac{D_a^2 k_B T}{8\pi^2 \hbar v_l^2 \rho} \frac{k}{E_k} \int \mathrm{d}\underline{q} \frac{1}{q} \delta\left(\frac{q}{2k} \pm \cos\theta'\right) \\ &= \frac{D_a^2 k_B T}{8\pi^2 \hbar v_l^2 \rho} \frac{k}{E_k} \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} q\, \mathrm{d}q \int_{-1}^{1} \mathrm{d}(\cos\theta')\, \delta\left(\frac{q}{2k} \pm \cos\theta'\right), \end{aligned} \tag{3.50}$$

where we have used the identity

$$\delta(ax) = \frac{1}{|a|} \delta(x). \tag{3.51}$$

Summing up absorption and emission processes (the two rates are equal in the elastic approximation), we have

$$W_{\mathrm{aco}}(\underline{k}) = \frac{D_a^2 k_B T}{8\pi^2 \hbar v_l^2 \rho} \frac{k}{E_{\underline{k}}} 2\pi \int_0^{2k} q\, \mathrm{d}q. \tag{3.52}$$

Notice that the integral over q has been restricted to the interval $[02k]$ since

$$\delta\left(\frac{q}{2k} \pm \cos\theta'\right) \to q = \mp 2k \cos\theta'. \tag{3.53}$$

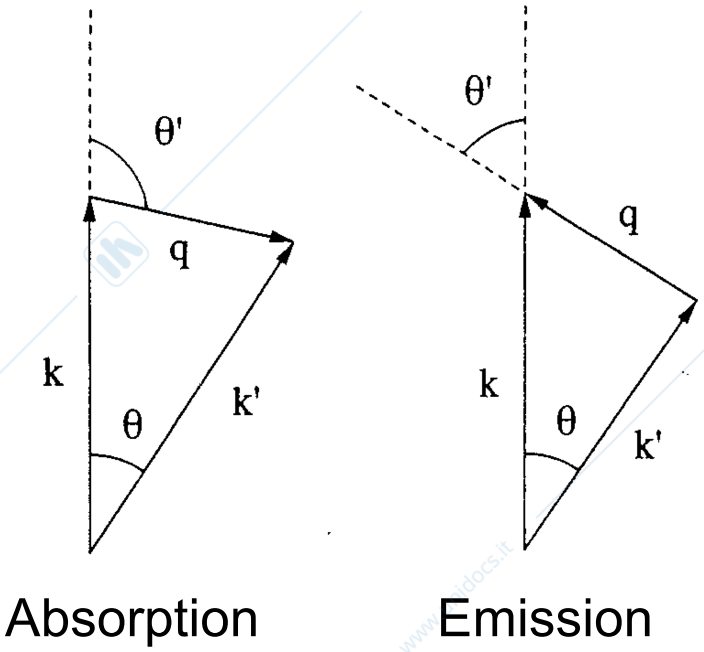


Figure 3.3.: Definition of angles θ and θ'. Absorption: $\underline{k}' = \underline{k} + \underline{q}$. Emission: $\underline{k}' = \underline{k} - \underline{q}$. Reproduced from [1].

The integral in (3.50) can be solved analytically, which results in the scattering rate [1, Eq. (2.73)]

$$W_{\text{aco}}(\underline{k}) = \frac{2\pi D_a^2 k_B T}{\hbar v_l^2 \rho} N(E_{\underline{k}}), \tag{3.54}$$

where

$$N(E_{\underline{k}}) = \frac{(2m^*)^{3/2}\sqrt{E_{\underline{k}}}}{4\pi^2\hbar^3} \tag{3.55}$$

is the density of states (without the spin factor) of a spherical parabolic valley. Eq. (3.54) is quite general, since it holds also for more realsitic band structure models, provided that the appropriate expression of the density of states is used. For values of k far from the minimum of the valley, the energy dispersion deviates from the simple parabolic expression $E_{\underline{k}} = \hbar^2 k^2/(2m^*)$, see Fig. 3.2, a better fit to the electronic structure can be obtained with the simple analytical expression

$$\gamma(E_{\underline{k}}) = E_{\underline{k}}(1 + \alpha E_{\underline{k}}) = \frac{\hbar^2 k^2}{2m^*} \tag{3.56}$$

which can be solved for $E_{\underline{k}}$

$$E_{\underline{k}} = \frac{2\gamma(E_{\underline{k}})}{1 + \sqrt{1 + 4\alpha\gamma(E_{\underline{k}})}}, \tag{3.57}$$

where α is the nonparabolicity coefficient. The corresponding density of states is

$$N(E_{\underline{k}}) = \frac{(2m^*)^{3/2}\sqrt{\gamma(E_{\underline{k}})}}{4\pi^2\hbar^3} \frac{\mathrm{d}\gamma(E_{\underline{k}})}{\mathrm{d}E_{\underline{k}}}. \tag{3.58}$$

The inelastic acoustic scattering rate (upper and lower expressions correspond to absorption and emission

processes, respectively) for a nonparabolic valley is [7, Eq. (2.4.30)]

$$W_{\mathrm{aco}}(E_{\mathrm{i}}) = \frac{(m^*)^{\frac{1}{2}}(k_{\mathrm{B}}T)^3 D_{\mathrm{a}}^2}{2^{\frac{5}{2}}\pi\hbar^4 v_l^4 \rho\sqrt{\gamma(E_{\mathrm{i}})}} \times \left\{ \begin{array}{l} (1+2\alpha E_{\mathrm{i}})[F_1(x_{2,\mathrm{a}}) - F_1(x_{1,\mathrm{a}})] + 2\alpha k_{\mathrm{B}}T[F_2(x_{2,\mathrm{a}}) - F_2(x_{1,\mathrm{a}})] \\ (1+2\alpha E_{\mathrm{i}})[G_1(x_{2,\mathrm{e}}) - G_1(x_{1,\mathrm{e}})] - 2\alpha k_{\mathrm{B}}T[G_2(x_{2,\mathrm{e}}) - G_2(x_{1,\mathrm{e}})] \end{array} \right\} \tag{3.59}$$

where we have defined the functions

$$F_1(x) = \int_0^x N_{\mathrm{q}}(z) z^2 \mathrm{d}z \tag{3.60a}$$

$$F_2(x) = \int_0^x N_{\mathrm{q}}(z) z^3 \mathrm{d}z \tag{3.60b}$$

$$G_1(x) = \int_0^x [N_{\mathrm{q}}(z)+1] z^2 \mathrm{d}z \tag{3.60c}$$

$$G_2(x) = \int_0^x [N_{\mathrm{q}}(z)+1] z^3 \mathrm{d}z \tag{3.60d}$$

and

$$N_q(z) = \frac{1}{e^z - 1} \tag{3.61}$$

is the Boson distribution function in terms of the dimensionless variable z. The integration limits in (3.60) are [7, Table 2.2]

$$\left\{ \begin{array}{l} x_{1,\mathrm{a}} = C(\alpha)\{\sqrt{E_{\mathrm{s}}}(1+2\alpha E_{\mathrm{i}}) - \sqrt{\gamma(E_{\mathrm{i}})}\} \\ x_{2,\mathrm{a}} = C(\alpha)\{\sqrt{E_{\mathrm{s}}}(1+2\alpha E_{\mathrm{i}}) + \sqrt{\gamma(E_{\mathrm{i}})}\} \end{array} \right. \qquad \gamma(E_{\mathrm{i}}) < \frac{E_{\mathrm{s}}}{1-4\alpha E_{\mathrm{s}}}$$

$$\left\{ \begin{array}{l} x_{1,\mathrm{a}} = 0 \\ x_{2,\mathrm{a}} = C(\alpha)\{\sqrt{\gamma(E_{\mathrm{i}})} + \sqrt{E_{\mathrm{s}}}(1+2\alpha E_{\mathrm{i}})\} \\ x_{1,\mathrm{e}} = 0 \\ x_{2,\mathrm{e}} = C(\alpha)\{\sqrt{\gamma(E_{\mathrm{i}})} - \sqrt{E_{\mathrm{s}}}(1+2\alpha E_{\mathrm{i}})\} \end{array} \right. \qquad \gamma(E_{\mathrm{i}}) > \frac{E_{\mathrm{s}}}{1-4\alpha E_{\mathrm{s}}}$$

with $E_{\mathrm{s}} = m^* v_l^2/2$, and $C(\alpha) = \frac{4\sqrt{E_{\mathrm{s}}}}{k_{\mathrm{B}}T(1-4\alpha E_{\mathrm{s}})}$.

Here are few Matlab lines necessary to define the relevant constants and material parameters that will be needed for the calculation of the scattering mechanisms in GaAs, see Table 3.4

```
c_light = 2.99792458e+8; % light velocity, m/s
H = 6.626070040e-34; % Planck constant, J*s
HBAR = 1.054D-34; % reduced Planck constant, J*s
kB = 1.3806488e-23; % Boltzmann constant, J/K
Q = 1.6021766208e-19; % elementary charge, C
T = 300; % temperature, K
eps0 = 8.854D-12; % vacuum permittivity constant, F/m
M0 = 9.1095D-31; % electron mass, kg
%
rho = 5320; % mass density, kg/m^3
eps_s = 12.9*eps0; % static dielectric constant, F/m
eps_infty = 10.9*eps0; % high-frequency dielectric constant, F/m
v_l = 5240; % longitudinal sound velocity, m/s
meff = 0.067*M0; % effective mass, kg
```

Properties	*Symbol*	*Si*	*GaAs*
Mass densities (kg/m^3)	ρ	2329.0	5320.0
Dielectric constant (F/m)	ϵ_S	11.7 ϵ_0	12.9 ϵ_0
	ϵ_∞	—	10.9 ϵ_0
Sound velocity (m/s)	v_S	9040	5240
Effective mass of electron	m^*/m_0	—	0.067 (Γ)
	m_l^*, m_t^*	0.92, 0.19 (X)	1.20, 0.2 (L)
Nonparabolicity factor	α	0.5 (X)	0.64 (Γ) 0.46 (L) 0.20 (X)
Longitudinal optical phonon energy (eV)	$\hbar\omega_0$	0.063	0.0354
Deformation potential:			
for acoustic phonon (eV)	Ξ_d	9.0	7.0
for optical phonon (eV/m)	D_{ij}	3.4×10^{10} (f_2)	1×10^{11} (Γ-L)
		4.0×10^{10} (f_3)	1×10^{11} (Γ-X)
		8.0×10^{9} (g_2)	1×10^{11} (L-L)
		3.0×10^{10} (g_3)	5×10^{10} (X-X)
Electron affinity (eV)	χ	4.05	4.07
Energy gap (eV)	E_g	1.12	1.424

Figure 3.4.: Material parameters for GaAs [1, Appendix A].

```
alpha = 0.64/Q; % nonparabolicity factor, 1/J
hwpop = 0.0354*Q; % longitudinal optical phonon energy, J
Daco = 7*Q; % acoustic deformation potential, J
egap = 1.424*Q; % energy gap, J
```

You can copy and paste these lines in your Matlab script. We strongly recommend the consistent use of units complying with the International System for all quantities. In all plots, energies should be conveniently expressed in eV. The integrals in (3.60) can be approximated by expanding the Bose-Einstein distribution function $N_q(z) = 1/[\exp(z) - 1]$ in terms of the dimensionless variable $z = (\hbar q v_l)/(k_B T)$, see Fig. 3.5

$$N_q(z) = \frac{1}{z} - \frac{1}{2} + \sum_{m=1}^{\infty} \frac{B_{2m} z^{2m-1}}{(2m)!}, \tag{3.62}$$

where B_{2m} are the Bernoulli numbers. Use the asymptotic limit

$$N_q(z) = e^{-z}, \tag{3.63}$$

for $z > \bar{z}$. Since there are some errors in [7, Appendix A], use Matlab's Symbolic Math Toolbox to derive the integrals in (3.60). Example:

```
B = [-1/2 1/6 0 -1/30 0 1/42 0 -1/30 0 5/66]; % Bernoulli numbers
% generated by bernoulli(sym(1:10))
```

$$F_1(x) = \begin{cases} \dfrac{x^2}{2} - \dfrac{x^3}{6} + \dfrac{x^4}{48} - \dfrac{x^6}{4320} + \dfrac{x^8}{241\,920} - \dfrac{x^{10}}{12\,096\,000} + \dfrac{x^{12}}{622\,702\,080}, & x \leqslant \bar{x}_a, \\ \dfrac{\bar{x}_a^2}{2} - \dfrac{\bar{x}_a^3}{6} + \dfrac{\bar{x}_a^4}{48} - \dfrac{\bar{x}_a^6}{4320} + \dfrac{\bar{x}_a^8}{241\,920} - \dfrac{\bar{x}_a^{10}}{12\,096\,000} + \dfrac{\bar{x}_a^{12}}{622\,702\,080} & \\ \quad + \exp(-\bar{x}_a)(\bar{x}_a^2 + 2\bar{x}_a + 2) - \exp(-x)(x^2 + 2x + 2), & x > \bar{x}_a, \end{cases} \tag{A.2}$$

$$G_1(x) = \begin{cases} \dfrac{x^2}{2} + \dfrac{x^3}{6} + \dfrac{x^6}{4320} + \dfrac{x^8}{241\,920} - \dfrac{x^{10}}{12\,096\,000} + \dfrac{x^{12}}{622\,702\,080}, & x \leqslant \bar{x}_e, \\ \dfrac{\bar{x}_e^2}{2} + \dfrac{\bar{x}_e^3}{6} + \dfrac{\bar{x}_e^4}{48} - \dfrac{\bar{x}_e^6}{4320} + \dfrac{\bar{x}_e^8}{241\,920} - \dfrac{\bar{x}_e^{10}}{12\,096\,000} + \dfrac{\bar{x}_e^{12}}{622\,702\,080} & \\ \quad + \exp(-\bar{x}_e)(\bar{x}_e^2 + 2\bar{x}_e + 2) - \dfrac{\bar{x}_e^3}{3} & \\ \quad - \exp(-x)(x^2 + 2x + 2) + \dfrac{x^3}{3}, & x > \bar{x}_e, \end{cases} \tag{A.3}$$

$$F_2(x) = \begin{cases} \dfrac{x^3}{3} - \dfrac{x^4}{8} + \dfrac{x^5}{60} - \dfrac{x^7}{5040} + \dfrac{x^9}{272\,160} - \dfrac{x^{11}}{143\,305\,600} + \dfrac{x^{13}}{622\,720\,080}, & x \leqslant \bar{x}_a, \\ \dfrac{\bar{x}_a^3}{3} - \dfrac{\bar{x}_a^4}{8} + \dfrac{\bar{x}_a^5}{60} - \dfrac{\bar{x}_a^7}{5040} + \dfrac{\bar{x}_a^9}{272\,160} - \dfrac{\bar{x}_a^{11}}{143\,305\,600} + \dfrac{\bar{x}_a^{13}}{622\,702\,080} & \\ \quad + \exp(-\bar{x}_a)(\bar{x}_a^3 + 3\bar{x}_a + 6\bar{x}_a + 6) & \\ \quad - \exp(-x)(x^3 + 3x^2 + 6x + 6), & x > \bar{x}_a, \end{cases} \tag{A.4}$$

$$G_2(x) = \begin{cases} \dfrac{x^3}{3} + \dfrac{x^4}{8} + \dfrac{x^5}{60} - \dfrac{x^7}{5040} + \dfrac{x^9}{272\,160} - \dfrac{x^{11}}{143\,305\,600} + \dfrac{x^{13}}{622\,702\,080}, & x \leqslant \bar{x}_e, \\ \dfrac{\bar{x}_e^3}{3} + \dfrac{\bar{x}_e^4}{8} + \dfrac{\bar{x}_e^5}{60} - \dfrac{\bar{x}_e^7}{5040} + \dfrac{\bar{x}_e^9}{272\,160} - \dfrac{\bar{x}_e^{11}}{143\,305\,600} + \dfrac{\bar{x}_e^{13}}{622\,702\,080} & \\ \quad + \exp(-\bar{x}_e)(\bar{x}_e^2 + 3\bar{x}_e^2 + 6\bar{x}_e + 6) - \dfrac{\bar{x}_e^4}{4} & \\ \quad - \exp(-x)(x^3 + 3x^2 + 6x + 6) + \dfrac{x^4}{4}, & x > \bar{x}_e, \end{cases} \tag{A.5}$$

Figure 3.5.: Numerical evaluation of some integrals of interest ($\bar{x}_a = 3.5$, $\bar{x}_e = 4$). Reproduced from [7, Appendix A].

```
z = sym('z','real');
x = sym('x','real');
nphon = 1/z - 1/2;
m=1; nphon = nphon + B(2*m)*z^(2*m-1)/factorial(2*m);
m=2; nphon = nphon + B(2*m)*z^(2*m-1)/factorial(2*m);
m=3; nphon = nphon + B(2*m)*z^(2*m-1)/factorial(2*m);
m=4; nphon = nphon + B(2*m)*z^(2*m-1)/factorial(2*m);
m=5; nphon = nphon + B(2*m)*z^(2*m-1)/factorial(2*m);
%
F1 = int(nphon*z^2,0,x);
F2 = int(nphon*z^3,0,x);
```

A Matlab function may be written in a separate file whose name defines the name of the function, with a filename extension of ".m". At the top of the file must be a line that contains the syntax definition for the new function, for example:

```
function [Waco_emi,Waco_abs] = aco_scat(E,T)
```

Having defined the function, the absorption and emission scattering rates may be plotted with the commands

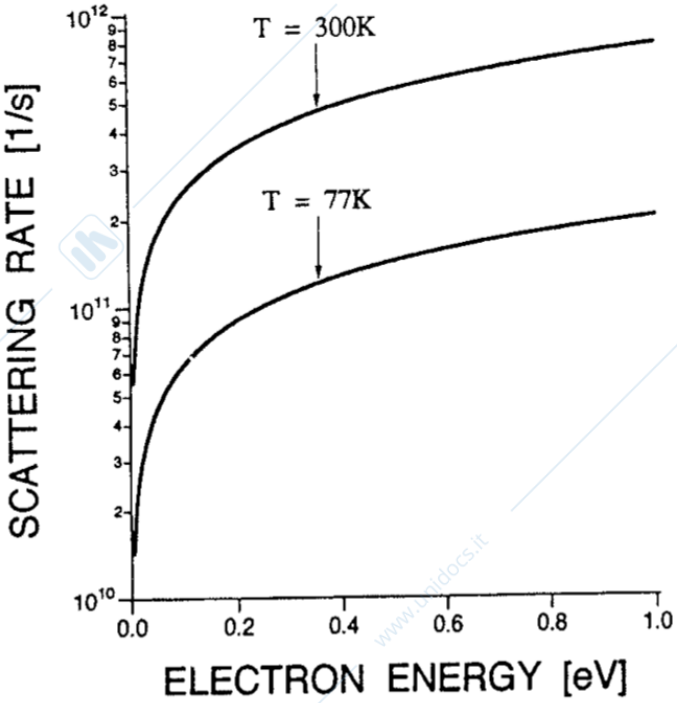


Figure 3.6.: Intravalley acoustic scattering in GaAs for electrons in the Γ-valley at different temperatures. [1, Fig. 2.10]

```
nE = 200; % number of energy points
vE = linspace(0,1,nE)*Q; % energy axis, J
Waco_abs = zeros(1,nE);
Waco_emi = zeros(1,nE);
for ie = 1:nE, E = vE(ie);
[Waco_emi(ie), Waco_abs(ie)] = aco_scat(E,T); end
%
figure(1), hold on
plot(vE/Q,Waco_emi,'r.-','linewidth',2)
plot(vE/Q,Waco_abs,'b.-','linewidth',2)
plot(vE/Q,Waco_emi+Waco_abs,'k.-','linewidth',2)
set(gca,'FontSize',14,'FontName','Arial','box','on')
ylabel('Acoustic scattering rate, 1/s), xlabel('Energy, eV')
legend('emission','absorption','total')
```

N.B.: functions for the calculation of the scattering rates are to be included in the single-particle Monte Carlo code described in the second part of this laboratory. As these functions are going to be called many times in a typical Monte Carlo run, they have to be efficient: please do not include any symbolic code here! Use the Symbolic Matlab Toolbox only to derive mathematical expressions.

Write a Matlab program to compute the acoustic deformation potential scattering rate in GaAs according to the inelastic expression (3.59); plot the absorption and emission scattering rates as a function of energy, for different values of the temperature (e.g., $T = 300\,\mathrm{K}$ and $T = 77\,\mathrm{K}$). Compare the results with the elastic approximation (3.54), see Fig. 3.6

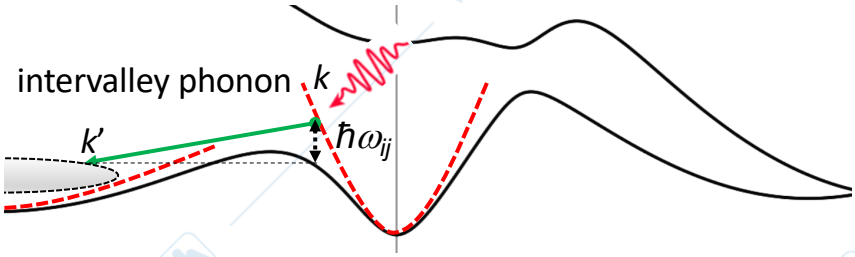
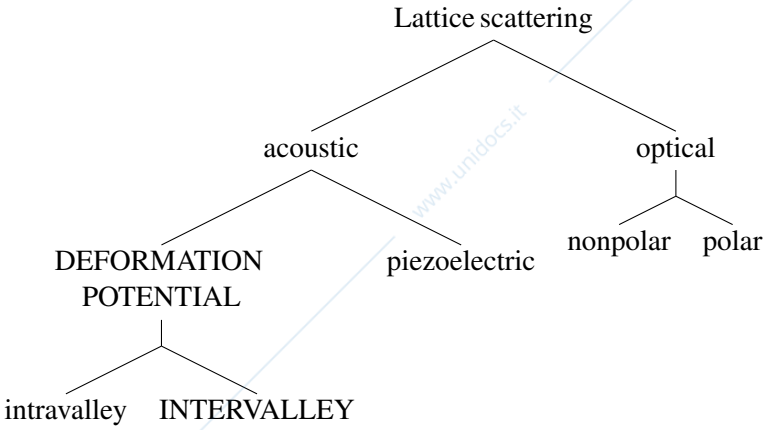


Figure 3.7.: Intervalley transition from the Γ-valley to the L-valley in GaAs. As opposed to the intravalley process in Fig. 3.1, the intervalley process cannot be approximated as elastic, because the wavevectors involved are large, see Fig. 3.3.



3.1.2. Deformation potential scattering by acoustic phonons (intervalley)

If the transition involves two different valleys, see Fig. 3.7, the phonon wavevector $\underline{q}$ is typically very close to the minima of the initial and final valleys (Fig. 3.7 could not be drawn in scale). Consequently, q is almost constant, and so is the intervalley phonon energy $\hbar\omega_{ij}$ involved in the scattering process connecting the initial valley i to the final valley j. The resulting intervalley scattering rate is similar to (3.54) [1, Eq. (2.86)]

$$W_{\mathrm{ij}}(E_{\underline{k}}) = \frac{\pi D_{ij}^2 Z_j}{\rho\,\omega_{ij}} \left(N_{ij} + \frac{1}{2} \mp \frac{1}{2} \right) N(\underbrace{E_{\underline{k}} \pm \hbar\omega_{ij} - \Delta E_{ji}}_{\mathrm{E}_{\underline{k}'}}), \tag{3.64}$$

where Z_j is the number of equivalent final valleys (see Fig. 3.9) for the considered intervalley process, N_{ij} is the number of intervalley phonons

$$N_{ij} = \frac{1}{e^{(\hbar\omega_{ij})/(k_B T)} - 1}, \tag{3.65}$$

and ΔE_{ji} is the energy difference between the minima of the final and initial valleys connected by the coupling constant D_{ij}, see [1, Appendix 1]. For valleys with ellipsoidal equienergetic surfaces, i.e., with different longitudinal and transverse components of the effective-mass tensor (e.g., the L valleys), the dispersion relation reads

$$E_{\underline{k}} = \frac{\hbar^2 k_l^2}{2m_l} + \frac{\hbar^2 k_t^2}{2m_t}, \tag{3.66}$$

while the density of states can be evaluated with the parabolic expression (3.55) with the DOS effective mass defined as

$$m^* = \left(m_l m_t^2\right)^{1/3}. \tag{3.67}$$

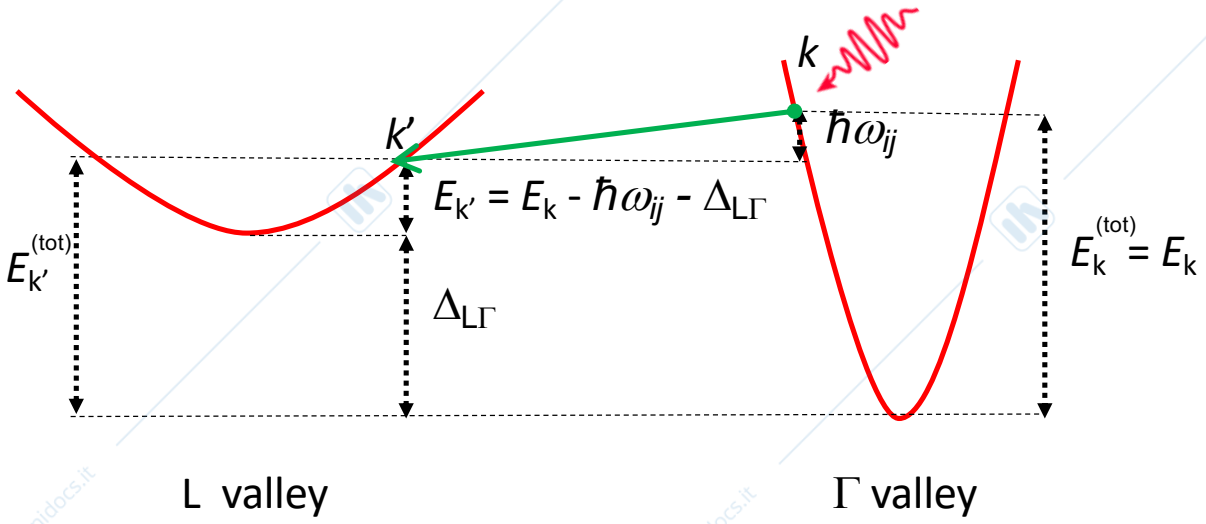


Figure 3.8.: An intervalley phonon emission process scattering an electron from the Γ valley to the L valley. In the calculation of the scattering rates, the initial and final energies $E_{\underline{k}}$ and $E_{\underline{k}'}$ appearing in the expression of the scattering rate (3.64) are the kinetic energy measured from the bottom of the corresponding valleys. N.B.: the density of states appearing in (3.64) is the density of states of the *final* valley.

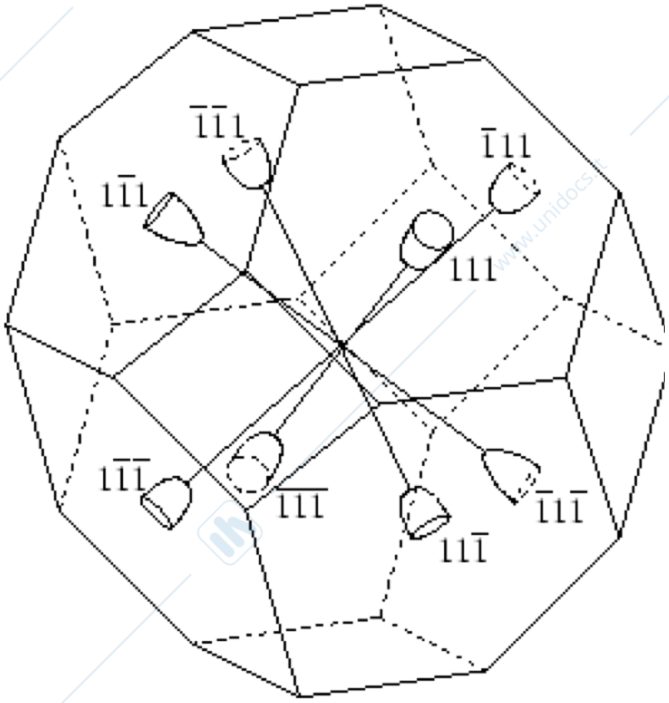


Figure 3.9.: The eight half-ellipsoids represent the constant energy surface associated to the L-valleys in GaAs. There are four equivalent L valleys in the Brillouin zone (each L-valley is equally shared by two Brillouin zones).

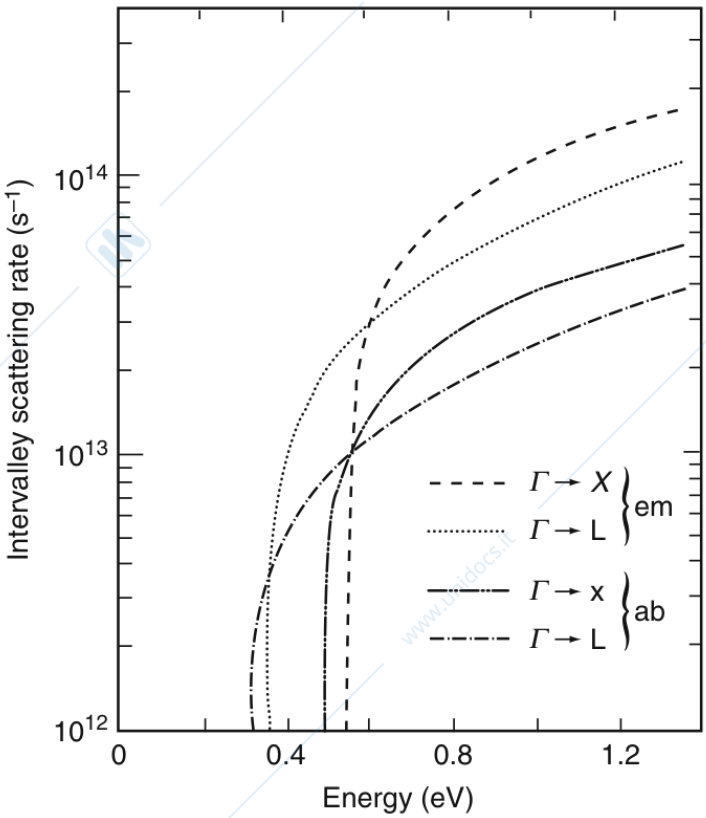


Figure 3.10.: Intervalley acoustic scattering in GaAs for electrons from the Γ-valley to the *L*- and *X*-valleys. The energy is measured from the bottom of the Γ valley [7, Fig. 2.15].

Compute the intervalley scattering rate (3.64) for the transition $\Gamma \to L$ in GaAs, see Fig. 3.10. See Fig. 3.4 for a list of the material parameters in GaAs. Assume an intervalley phonon energy $\hbar\omega_{ij} = 28\,\text{meV}$ for Γ-L transitions [19, Table 7.2].

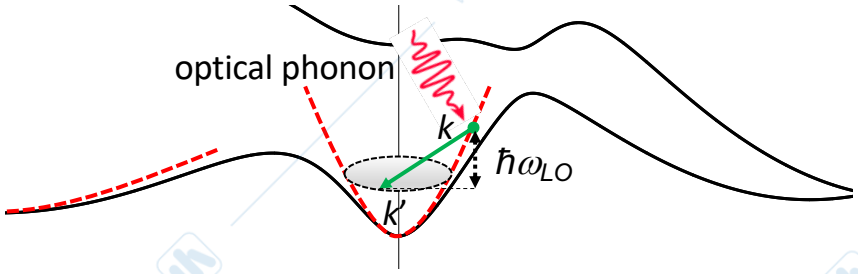
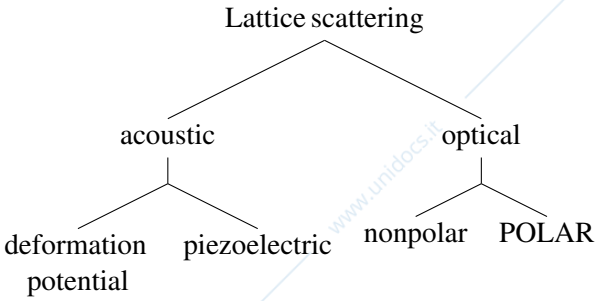


Figure 3.11.: Intravalley transition assisted by a polar optical phonon. Only transitions that conserve energy and momentum are allowed.



3.1.3. Polar scattering by optical phonons

In addition to deformation potential scattering, the vibrations of opposite charged atoms in polar materials generate long-range dipole fields giving rise to a macroscopic electric field that can scatter the carriers. The polar interaction may be due to either acoustic or optical phonons, leading to piezoelectric and polar optical scattering, respectively. The latter is the dominant scattering process in III-V and II-VI compound semiconductors. Scattering with polar optical phonons can be described by the Fröhlich formulation. Let's first recall that only longitudinal optical (LO) modes give rise to a dipole field. Assuming a dispersionless LO phonons, the dipole field is proportional to the relative displacement of the two atoms in the unit cell [2, Chapter 3]

$$\hat{\underline{\Pi}}(\underline{r}) = \sqrt{\frac{\hbar\omega_{\mathrm{LO}}}{2\Omega}\left(\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_s}\right)}\sum_{\underline{q}}\left(\hat{a}_{\underline{q}}e^{\mathrm{i}\underline{q}\cdot\underline{r}}+\hat{a}^\dagger_{\underline{q}}e^{-\mathrm{i}\underline{q}\cdot\underline{r}}\right)\underline{e}_{\underline{q}}, \tag{3.68}$$

where ε_∞, ε_s are the optical and static dielectric functions (the values for GaAs are reported in Fig. 3.4). The difference $\varepsilon_s - \varepsilon_\infty$ represents the contribution of the ions to the dielectric response[4] and it is also known as the *oscillator strength* or *f-strength*. The polarization charge of the lattice is nonzero only for longitudinal fields ($\underline{e}_{\underline{q}} = \underline{q}/q$)

$$\hat{\rho}_p(\underline{r}) = -\nabla\cdot\hat{\underline{\Pi}}(\underline{r}) = -\mathrm{i}\sqrt{\frac{\hbar\omega_{\mathrm{LO}}}{2\Omega}\left(\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_s}\right)}\sum_{\underline{q}} q\left(\hat{a}_{\underline{q}}e^{\mathrm{i}\underline{q}\cdot\underline{r}}-\hat{a}^\dagger_{\underline{q}}e^{-\mathrm{i}\underline{q}\cdot\underline{r}}\right). \tag{3.69}$$

[4] Any external perturbation causes not only the single-particle scattering process we are interested in, but also a many-body rearrangement of all electrons, valence and conduction, as well as of the ions (especially if the crystal is ionic) [4, Chapter 17]. Thus, we must consider both the electronic and the ionic dielectric properties. This rearrangement modifies the scattering potential, so that the scattering event depends on these many-body effects as well. In introductory textbooks of electromagnetic theory, the effect is lumped into a dielectric constant of the medium. That would be too much of a simplification in our case, as in general the response of the system depends on both the wavevector $\underline{q}$ and the frequency ω of the external perturbation, which, without loss of generality, we can assume to be of the harmonic form $e^{\mathrm{i}\omega t}e^{\mathrm{i}\underline{q}\cdot\underline{r}}$. In the literature, ε_∞ represents the dielectric constant in absence of the ionic response, i.e., at frequencies large enough so that the ions cannot respond. As the frequency decreases, the ions begin to respond and their response results in a larger dielectric constant ε_s. The response of the conduction electrons is called *free carrier screening*, and is usually described by means of the Debye-Hückel approximation (3.71).

Therefore, the expression for the perturbation Hamiltonian associated with the electrostatic interaction is (3.5)

$$\hat{H}' = -e\int \mathrm{d}\underline{r}' \underbrace{\frac{e^{-q_0|\underline{r}-\underline{r}'|}}{4\pi|\underline{r}-\underline{r}'|}}_{\mathrm{G}(\underline{r},\underline{r}')} \hat{\rho}_p(\underline{r}') = \mathrm{i}\sum_{\underline{q}} g_{\underline{q}} \left(\hat{a}_{\underline{q}} e^{\mathrm{i}\underline{q}\cdot\underline{r}} - \hat{a}^{\dagger}_{\underline{q}} e^{-\mathrm{i}\underline{q}\cdot\underline{r}}\right) \tag{3.70}$$

where $G(\underline{r},\underline{r}')$ is the Green's function of the Poisson equation (the screened Coulomb potential due to an impusilve charge at $\underline{r}'$), and

$$q_0 = \sqrt{e^2 n/(\varepsilon_s k_B T)} \tag{3.71}$$

is the reciprocal screening length (Debye-Hückel screening in the static long-wavelength limit for non-degenerate statistics [20, Chapter 8]), n is the electron density, and

$$g_{\underline{q}} = \sqrt{\frac{e^2\hbar\omega_{LO}}{2\Omega}\left(\frac{q}{q^2+q_0^2}\right)^2\left(\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_s}\right)}. \tag{3.72}$$

Application of Fermi's Golden rules gives the scattering probability

$$S_{\mathrm{pop}}(\underline{k},\underline{k}') = \frac{2\pi}{\hbar}\frac{e^2\hbar\omega_{\mathrm{LO}}}{2\Omega\varepsilon_p}\left\{N_q+\frac{1}{2}\mp\frac{1}{2}\right\}\left(\frac{q}{q^2+q_0^2}\right)^2 \delta[E(\underline{k}')-E(\underline{k})\mp\hbar\omega_{\mathrm{LO}}], \tag{3.73}$$

with $\varepsilon_p = 1/(1/\varepsilon_\infty - 1/\varepsilon_\mathrm{s})$. The dependence of the matrix element on the inverse square of the phonon wavevector, and the angular dependence of the overlap integral makes this scattering mechanism anisotropic. Integration over the final states gives for a parabolic valley with $E(\underline{k}) = \hbar^2k^2/(2m^*)$

$$\begin{aligned} W_{\mathrm{pop}}(\underline{k}) &= \frac{2\pi}{\hbar}\frac{e^2\hbar\omega_{\mathrm{LO}}}{2\Omega\varepsilon_p}\frac{\Omega}{(2\pi)^3}\int_{\mathrm{BZ}}\mathrm{d}\underline{k}'\left\{N_q+\frac{1}{2}\mp\frac{1}{2}\right\}\left(\frac{q}{q^2+q_0^2}\right)^2\delta[E(\underline{k}')-E(\underline{k})\mp\hbar\omega_{\mathrm{LO}}] \\ &= \frac{e^2\omega_{\mathrm{LO}}}{8\pi^2\varepsilon_p}\int_{\mathrm{BZ}}\mathrm{d}\underline{q}\left\{N_q+\frac{1}{2}\mp\frac{1}{2}\right\}\left(\frac{q}{q^2+q_0^2}\right)^2\delta\left(\frac{\hbar^2q^2}{2m^*}\pm\frac{\hbar^2kq\cos\theta'}{m^*}\mp\hbar\omega_{\mathrm{LO}}\right). \end{aligned} \tag{3.74}$$

Neglecting screening ($q_0 = 0$)

$$W_{\mathrm{pop}}(\underline{k}) = \frac{e^2\omega_{\mathrm{LO}}}{8\pi^2\varepsilon_p}\int_0^{2\pi}\mathrm{d}\phi\int_{-1}^{1}\mathrm{d}(\cos\theta')\int_0^{\infty}q^2\mathrm{d}q\left\{N_q+\frac{1}{2}\pm\frac{1}{2}\right\}\frac{1}{q^2}\delta\left(\frac{\hbar^2q^2}{2m^*}\pm\frac{\hbar^2kq\cos\theta'}{m^*}\mp\hbar\omega_{\mathrm{LO}}\right), \tag{3.75}$$

which gives the energy-dependent scattering rate [1, Eq. (2.100)]

$$W_{\mathrm{pop}}(E_{\underline{k}}) = \frac{e^2\omega_{\mathrm{LO}}}{8\pi\varepsilon_p}\frac{k}{E_{\underline{k}}}\left\{N_q+\frac{1}{2}\mp\frac{1}{2}\right\}\ln\left(\frac{q_{\max}}{q_{\min}}\right), \tag{3.76}$$

with $q_{\min} = k\left|1-\sqrt{1\pm\frac{\hbar\omega_{\mathrm{LO}}}{E_{\underline{k}}}}\right|$, and $q_{\max} = k\left|1+\sqrt{1\pm\frac{\hbar\omega_{\mathrm{LO}}}{E_{\underline{k}}}}\right|$. These integration interval $[q_{\min}, q_{\max}]$ is imposed by

$$\delta\left(\frac{\hbar^2q^2}{2m^*}\pm\frac{\hbar^2kq\cos\theta'}{m^*}\mp\hbar\omega_{\mathrm{LO}}\right) = \frac{m^*}{\hbar^2kq}\delta\left(\frac{q}{2k}\pm\cos\theta'\mp\frac{k\hbar\omega_{\mathrm{LO}}}{2qE_{\underline{k}}}\right),$$

which gives the integration limits since $\cos\theta'$ is limited in the interval $[-1,1]$.

For a nonparabolic band with $\gamma(E_{\underline{k}}) = E_{\underline{k}}(1+\alpha E_{\underline{k}})$, see (3.56), the scattering rate is [7, Eq. (2.4.57)]

$$W_{\mathrm{pop}}(E_{\mathrm{i}}) = \frac{e^2\sqrt{m^*}\omega_{\mathrm{LO}}}{4\pi\varepsilon_p\sqrt{2}\hbar\sqrt{\gamma(E_{\mathrm{i}})}}\frac{\partial\gamma(E_{\mathrm{f}})}{\partial E_{\mathrm{f}}}\left\{N_q+\frac{1}{2}\mp\frac{1}{2}\right\}\frac{1}{C}\left\{A\ln\left|\frac{\gamma^{\frac{1}{2}}(E_{\mathrm{i}})+\gamma^{\frac{1}{2}}(E_{\mathrm{f}})}{\gamma^{\frac{1}{2}}(E_{\mathrm{i}})-\gamma^{\frac{1}{2}}(E_{\mathrm{f}})}\right|+B\right\} \tag{3.77}$$

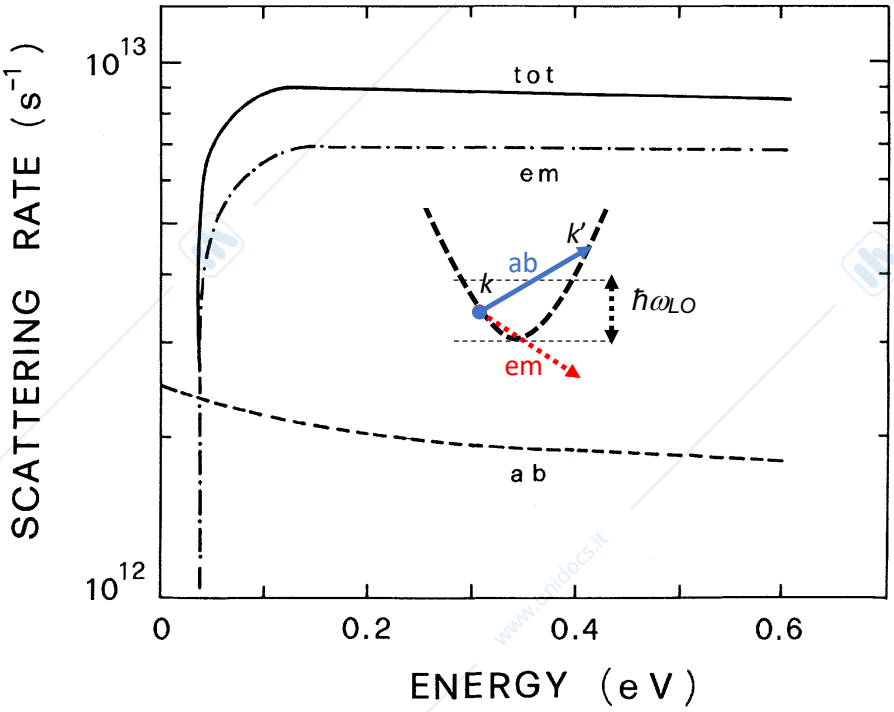


Figure 3.12.: Polar optical scattering rate for the Γ valley in bulk GaAs at $T = 300\,\mathrm{K}$. Reproduced from [7, Fig. 2.17]. N.B. that the emission scattering rate is zero when $E < \hbar\omega_{\mathrm{LO}}$, because no final states are available below the conduction band edge, see inset.

with

$$A = \{2(1+\alpha E_{\mathrm{i}})(1+\alpha E_{\mathrm{f}}) + \alpha[\gamma(E_{\mathrm{i}}) + \gamma(E_{\mathrm{f}})]\}^2 \tag{3.78a}$$

$$B = -2\alpha\gamma^{\frac{1}{2}}(E_{\mathrm{i}})\gamma^{\frac{1}{2}}(E_{\mathrm{f}})\{4(1+\alpha E_{\mathrm{i}})(1+\alpha E_{\mathrm{f}}) + \alpha[\gamma(E_{\mathrm{i}}) + \gamma(E_{\mathrm{f}})]\} \tag{3.78b}$$

$$C = 4(1+2\alpha E_{\mathrm{i}})(1+2\alpha E_{\mathrm{f}})(1+\alpha E_{\mathrm{i}})(1+\alpha E_{\mathrm{f}}) \tag{3.78c}$$

where E_{i} and $E_{\mathrm{f}} = E_{\mathrm{i}} \pm \hbar\omega_{\mathrm{LO}}$ are the initial and final energies before and after the collision.

Compute the polar optical scattering rate in GaAs according to the inelastic expression (3.77). Plot emission and absorption rates as a function of energy, see Fig. 3.12. Compare the results with the parabolic approximation (3.76).

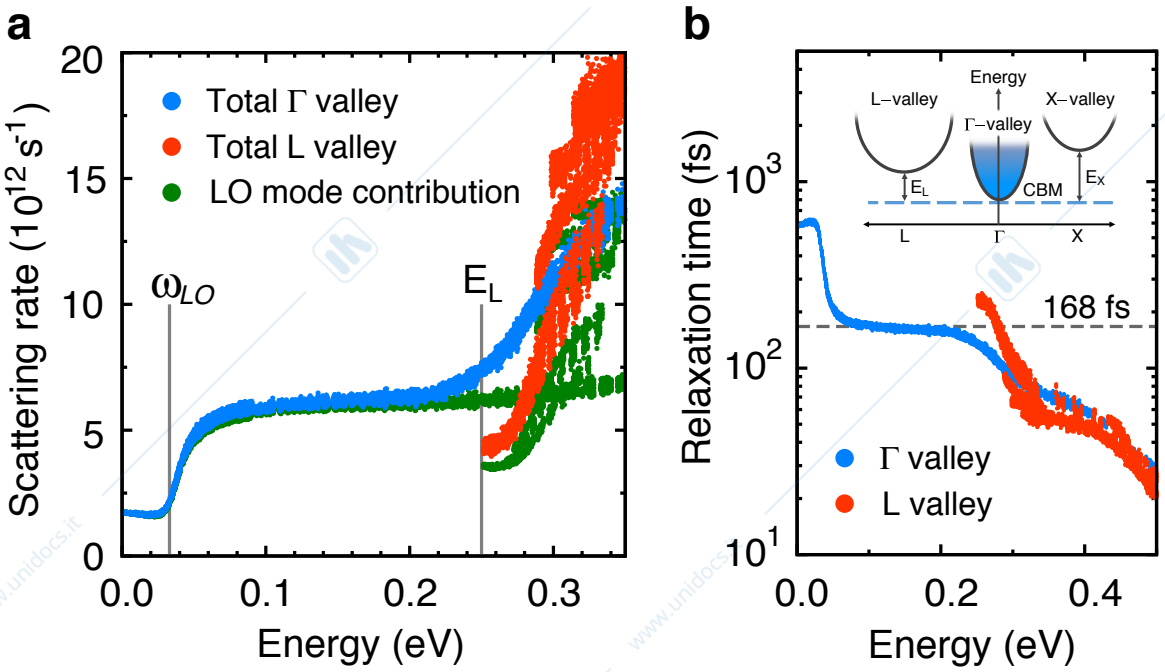


Figure 3.13.: (a) Electron-phonon (e-ph) scattering rates and (b) the corresponding relaxation times (the inverse of the scattering rates) computed at $T = 300$ K for electrons in GaAs with energies within 0.4 eV of the conduction band minimum, which is the origin of the energy axis. Data points in blue (red) are for electronic states in the Γ (L) valley. The scattering rate associated with LO phonon scattering alone is also shown in (a). Reproduced from [21].

3.1.4. Comparison with first-principles calculations and experiments

We are now ready to draw some conclusions concerning carrier-phonon (e-ph) scattering. Absent in covalent materials like Si and Ge (for which the absence of ionic polarization implies $\varepsilon_s = \varepsilon_\infty$), Fröhlich scattering is the dominant intravalley scattering process affecting the low-field mobility of III-V compound semiconductors, while acoustic deformation potential interaction dominates intervalley scattering and hot carrier dynamics.[5] It is interesting to compare our results with first-principles calculations. Fig.3.13 shows e-ph scattering rate (deformation potential plus polar interaction) and the corresponding relaxation times in GaAs computed with density functional theory (DFT) [21]. For electrons in the Γ valley, only intravalley scattering is possible for energies up to E_L. Small-q LO phonon scattering (blue dots) dominates in this energy range. The scattering rate is nearly constant over the 0.05-0.25 eV energy range, with an associated relaxation time of ≈ 168 fs. At energies below ≈ 0.05 eV the scattering rate drops sharply because the phase space for LO phonon emission vanishes, and the scattering process is dominated by LO phonon absorption. In this energy range the relaxation time is found to be close to 600 fs. At energy higher than E_L, Γ-L intervalley scattering becomes possible, and the scattering rate increases rapidly as a result. Intravalley scattering in the L valley, also possible above E_L, is dominated by polar phonons. It exhibits a scattering rate with multiple branches, and thus a strong k-dependence, due to the anisotropy of the L valley. For energies above E_L, the $\Gamma - L$ intervalley scattering is dominated by large-q LA and TA acoustic phonon scattering.

[5] As long as the electric field is low enough, polar optical scattering keeps the electrons at relatively low kinetic energies. In this low-energy range the scattering rates $W_{\text{pop}}(E)$ increases with increasing E. If the electric field increases above some critical value, electrons are accelerated to the range of higher kinetic energies in which $W_{\text{pop}}(E)$ decreases with increasing E, and so they scatter less frequently, thus losing a smaller fraction of their kinetic energy to LO-phonons, and, in the absence of other inelastic scattering mechanisms, gaining from the field even more kinetic energy, etc. This diverging process is called electron "run-away".

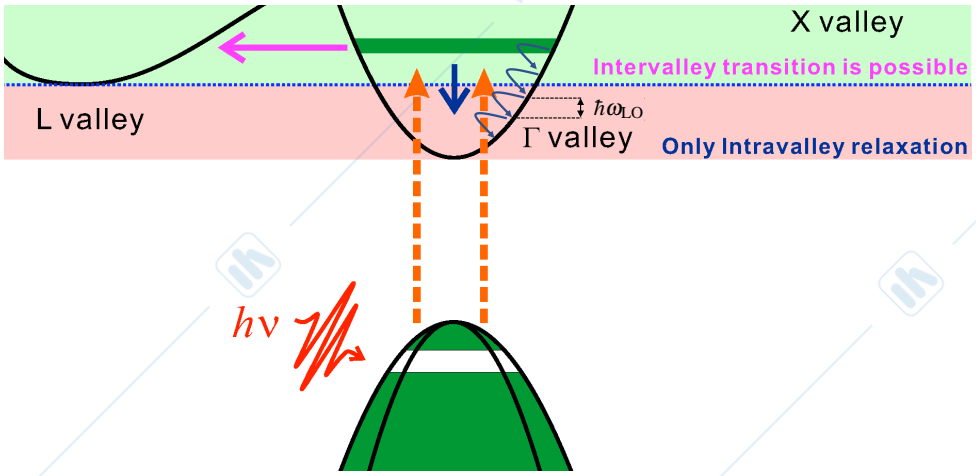


Figure 3.14.: Band structure excitation and relaxation scheme in GaAs for the ultrafast pump-probe experiments on quantum kinetic electron-phonon scattering: Electrons are generated in the conduction band by an ultrashort laser pulse, leaving holes in the valence band. The photogenerated electrons relax via emission of longitudinal optical (LO) phonons of energy $\hbar\omega_{LO}$.

The scattering rates are not directly accessible to experiments. However, information about the relaxation times can be obtained from time-resolved experiments of carrier relaxation times, which describe how optically excited carriers relax from a monoenergetic distribution to a quasi-equilibrium distribution at the band edges, see Fig. 3.14. In general, electron-electron scattering may play an important role in the initial momentum relaxation process, in which the quasi monoenergetic distribution relaxes to a *hot* Fermi distribution with a temperature higher than the lattice temperature. It is only after this ultrafast thermalization process is completed (few femptoseconds after photoexcitation), that this hot carrier distribution loses energy by interacting with the lattice. It is generally understood that the hot electrons lose their energy mainly to polar optical phonons, but recent studies indicate that intervalley scattering plays an important role if the electrons are photogerenated at sufficiently high energies, see Fig. 3.15 and 3.16. A similar argument applies also to silicon according to recent full-band Monte Carlo simulations by Max Fischetti, see excerpt below [22]

most of the electron energy, but no disagreement seems to arise on the fact that emission of optical phonons is indeed the major cause of energy loss.

This is such a well-known fact that seldom, if ever, is its validity questioned and seldom, if ever, are references provided. The history of this "truth" is indeed quite old, but calculations we have performed in the past seem to indicate that this statement may be more myth than truth. Clearly, if electrons were to lose energy mostly to acoustic phonons, the power-dissipation issue should be revisited. Since phonons would be generated over a rather large distribution of wavelengths, two competing effects would contribute to increasing or decreasing heat conduction, depending on the group velocity of the excited phonons: On the one hand, long-wavelength acoustic phonons would diffuse quickly and heat conduction would occur without having to wait for the decay of optical phonons into acoustic excitations with a higher group velocity. This would result in faster heat conduction. On the other hand, since short-wavelength (zone-edge) acoustic phonons also move at a small group velocity, similar to optical modes, one may be tempted to lump all zone-edge modes under the single label of "optical phonons," as indeed was done in the past.[15] However, this would be

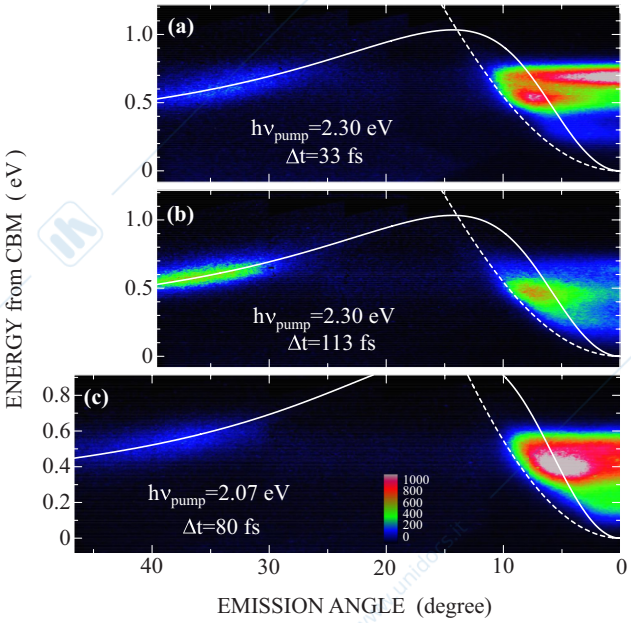


Figure 3.15.: The photoemission map for hot electrons injected into the conduction band with $h\nu_{\text{pump}} = 2.30$ eV light pulses, at 33 (a) and 113 (b) fs after excitation. One can see that at $\Delta t = 113$ fs, a significant part of electronic packet is already transferred to the eight L valleys (of which only one is shown), and also, that the hot electrons have lost energy with respect to the distribution shown in (a). For a lower pump $h\nu_{\text{pump}} = 2.07$ eV (c), the photoemission map at 80 fs after excitation shows that the transfer to the L valley is slower and the electrons transferred to the L valley are located at lower energy than those in (a). Measurements were made at 90 K. The solid and broken curves show the dispersion along L and X directions. The color scale indicates the photoemission intensity. At low concentrations, the role of electron-electron scattering can be minimized, and the hot electron relaxation can be entirely ascribed to carrier-phonon scattering. Reproduced from [23].

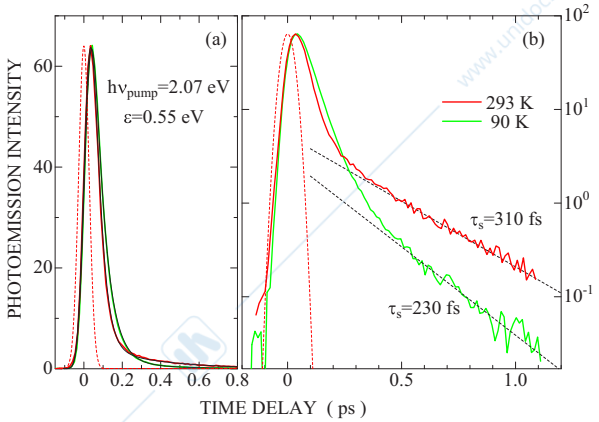
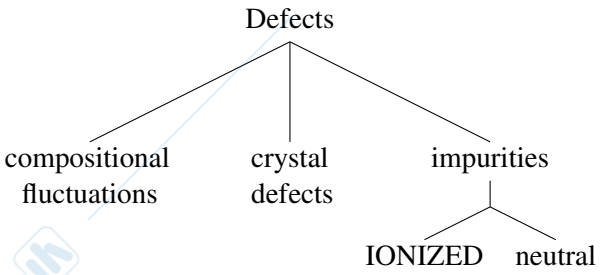


Figure 3.16.: Temporal changes in hot-electron populations with excess energy of 0.55 eV in the Γ valley, measured at 90 and 293 K. The photoemission intensities which represent the populations are plotted on a linear scale in panel (a) and on a semilogarithmic scale in panel (b). The broken lines in panel (b) show the decay of slow components with time constants of 230 and 310 fs at 90 and 293 K, respectively. There is clearly a bimodal decay of the population at both temperatures; an initial fast decay due to momentum relaxation is followed by a slower energy relaxation. Reproduced from [23].



3.2. Ionized impurity scattering (Brooks and Herring)

Phonon scattering is an intrinsic process, since it is present in a perfect crystal. Carriers in semiconductor devices are usually supplied/removed through heavily doped regions, which may be regarded as carrier reservoirs. In such doped regions, carrier transport is substantially affected by scattering due to randomly distributed impurities, which are ionized by design, so that the carriers resulting from ionization can contribute to carrier transport. Each impurity is strongly bonded to the whole lattice, so that the momentum transferred to the lattice is negligible. If $V(\underline{r}-\underline{r}_i)$ is the potential of the impurity at $\underline{r}_i$, the Hamiltonian describing the electron-impurity interaction is

$$H'(\underline{r}) = \sum_{i=1}^{N_I} V(\underline{r}-\underline{r}_i), \tag{3.79}$$

where N_I is the total number of impurities. Assuming *Normal* processes and ignoring the overlap factor, the matrix element is (H' does not act on the crystal variables)

$$\begin{aligned}\langle \underline{k}'|H'|\underline{k}\rangle = \frac{1}{\Omega}\int_\Omega \mathrm{d}\underline{r} H'(\underline{r}) e^{\mathrm{i}(\underline{k}'-\underline{k})\cdot\underline{r}} = \\ \frac{1}{\Omega}\int_\Omega \mathrm{d}\underline{r} \sum_{i=1}^{N_I} V(\underline{r}-\underline{r}_i) e^{\mathrm{i}(\underline{k}'-\underline{k})\cdot\underline{r}} = \\ V_{\underline{k}-\underline{k}'} \sum_i e^{-\mathrm{i}(\underline{k}-\underline{k}')\cdot\underline{r}_i},\end{aligned} \tag{3.80}$$

where $V_{\underline{k}}$ is the Furier transform of the potential due to the single impurity in the orign

$$V_{\underline{k}} = \mathscr{F}\{V(\underline{r})\} = \frac{1}{\Omega}\int_\Omega \mathrm{d}\underline{r} V(\underline{r}) e^{-\mathrm{i}\underline{k}\cdot\underline{r}}, \tag{3.81}$$

and we have used the property of the Fourier transform

$$\mathscr{F}\{V(\underline{r}-\underline{r}_i)\} = V_{\underline{k}} e^{-\mathrm{i}\underline{k}\cdot\underline{r}_i}. \tag{3.82}$$

In principle, we could solve the exact Schrödinger equation in the presence of the impurities, by including the Hamiltonian (3.79) in the unperturbed Hamiltonian and stop talking about impurity scattering altogether. The effect of the impurities on carrier transport would be that of modifying the wavefunctions, and so the transmission amplitude across the sample, via the coherent interaction between the electron wavefunction and the impurity potential. However, in writing (3.79), we have implicitly assumed that the positions of all the impurities are precisely known. In experimental situations this is obviously not the case. One must regard simulations of charge transport in a given sample as an ensemble average over many configurations, each of them corresponding to a different set of random (but still yielding the same average impurity density) impurity positions $\underline{r}_i$. In doing this average, all interference among partial waves scattered from different impurities is lost (or averaged out), as in the usual random phase approximation (RPA). Assuming a spatially uniform distribution of impurities, we find the average expression:

$$\left|\langle \underline{k}'|H'|\underline{k}\rangle\right|^2 = \left|V_{\underline{k}-\underline{k}'}\right|^2 \sum_{ii'} \underbrace{e^{-\mathrm{i}(\underline{k}-\underline{k}')\cdot(\underline{r}_i-\underline{r}_{i'})}}_{\approx\delta_{ii'}} \approx \left|V_{\underline{k}-\underline{k}'}\right|^2 N_I. \tag{3.83}$$

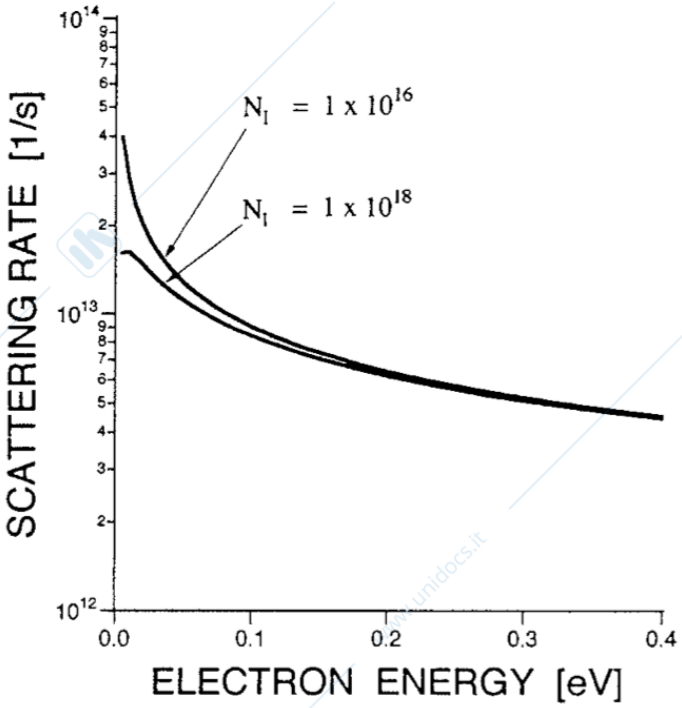


Figure 3.17.: Impurity scattering computed for different values of the doping concentration (impurities are considered fully ionized). [1, Fig. 2.5].

The scattering potential $V(\underline{r})$, cannot be taken as the bare Coulomb potential of the ionized impurity, as both valence and conduction electrons, when present in significant concentrations, will screen the potential. This actually helps, since the cross section for scattering with a bare Coulomb potential diverges as a result of the infinite range of the potential. Moreover, when the impurities are sufficiently diluted, we can consider them as isolated and screen them one at a time with the static free-carrier dielectric function, which causes the potential to drop off more rapidly with r than a purely Coulombic potential inversely proportional to r[6]

$$V(r) = \frac{Ze^2}{4\pi\varepsilon_s r} e^{-q_0 r} \tag{3.84}$$

where Z is the number of charge units of the impurity (typically $Z = 1$), and q_0 is the reciprocal screening length. The Fourier transform of (3.84) is

$$V_{\underline{k}-\underline{k}'} = \frac{Ze^2}{\varepsilon_s \Omega(|\underline{k}-\underline{k}'|^2 + q_0^2)}, \tag{3.85}$$

[6] At large impurity densities, when the average distance between impurities is smaller than the electron wavelength, the electrons are not able to screen the potentail, and the BH approach fails. At higher impurity concentrations additional problems arise. Impurities may cluster as the solid-solubility limit is reached, which is the case in Si for impurity concentrations well above $10^{20}/\text{cm}^3$. The dopant atoms are also close enough for the bound states in the impurity potential to delocalize, and electrons can tunnel among donor impurities, thus forming impurity bands that distort the band structure near the edge of the conduction band (or the valence band, for acceptors). When a high degree of compensation is present, very few free carriers are available to screen a much larger number of ionized (positive and negative) impurities. The latter, however, effectively screen each other, so that the Conwell and Weisskopf (CW) approach, which assumes a bare potential for $r \leq b$, and a cut-off of the potential is assumed at the mean distance b between the impurities, seems to be more appropriate for this condition. An attempt to reconcile BH and CW approaches is presented in [24, Chapter 4].

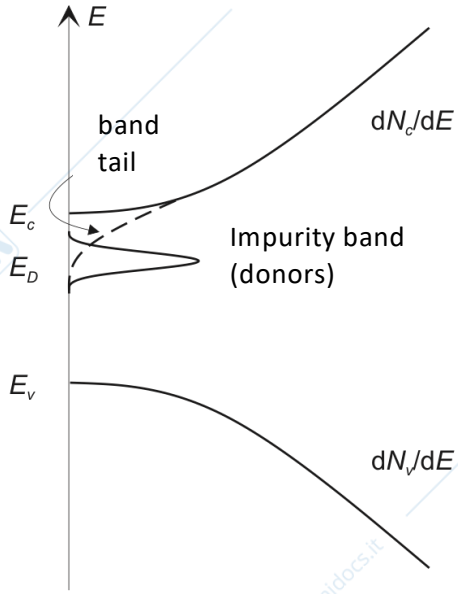


Figure 3.18.: As the impurity density becomes large, the interaction between adjacent impurity atoms leads to the splitting of the impurity levels into an impurity band. The interaction between free carriers and impurities results in the formation of band tails (band gap narrowing). Electrons can propagate within the impurity band, a trasport process known as *impurity band conduction*. [25, Fig. 3.7].

and we are now ready to compute the scattering rate. Considering that the Hamiltonian does not involve the crystal coordinates, we find the scattering probability

$$S_{\text{imp}}^{(\text{BH})}(\underline{k},\underline{k}') = \frac{2\pi}{\hbar} N_I \left|V_{\underline{k}-\underline{k}'}\right|^2 \delta(E_{\underline{k}'} - E_{\underline{k}}) \tag{3.86a}$$

$$= \frac{2\pi}{\hbar} N_I \frac{Z^2 e^4}{\varepsilon_s^2 \Omega^2 (|\underline{k}-\underline{k}'|^2 + q_0^2)^2} \delta(E_{\underline{k}'} - E_{\underline{k}}) \tag{3.86b}$$

$$= \frac{2\pi}{\hbar} n_I \frac{Z^2 e^4}{\varepsilon_s^2 \Omega [2k^2(1-\cos\theta) + q_0^2]^2} \delta(E_{\underline{k}'} - E_{\underline{k}}) \tag{3.86c}$$

where we have introduced the impurity density $n_I = N_I/\Omega$, and θ is the angle between $\underline{k}$ and $\underline{k}'$. Integration over final states gives [1, Eq. (2.49)]

$$W_{\text{imp}}^{(\text{BH})}(\underline{k}) = \frac{2\pi}{\hbar} n_I \frac{Z^2 e^4}{\varepsilon_s^2 \Omega} \frac{\Omega}{(2\pi)^3} \int_{\text{BZ}} \mathrm{d}\underline{k}' \frac{\delta(E_{\underline{k}'} - E_{\underline{k}})}{[2k^2(1-\cos\theta) + q_0^2]^2} \tag{3.87a}$$

$$= \frac{2\pi}{\hbar} n_I \frac{Z^2 e^4}{\varepsilon_s^2 \Omega} \frac{\Omega}{(2\pi)^3} \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \int_0^{\infty} \mathrm{d}k' \frac{k'^2 \sin\theta\, \delta(E_{\underline{k}'} - E_{\underline{k}})}{[2k^2(1-\cos\theta) + q_0^2]^2} \tag{3.87b}$$

$$= \frac{\pi n_I Z^2 e^4 N(E_{\underline{k}})}{\hbar \varepsilon_s^2} \int_{-1}^{1} \frac{\mathrm{d}(\cos\theta)}{[2k^2(1-\cos\theta) + q_0^2]^2}, \tag{3.87c}$$

which gives

$$W_{\text{imp}}^{(\text{BH})}(\underline{k}) = \frac{\pi n_I Z^2 e^4 N(E_{\underline{k}})}{\hbar \varepsilon_s^2} \frac{2}{q_0^2(4k^2 + q_0^2)}, \tag{3.88}$$

where $N(E_{\underline{k}})$ is the density of states, as defined in (3.55) for parabolic bands, and in (3.58) for non-parabolic bands.

Compute the impurity scattering rate in GaAs according to expression (3.88), for different values of the doping density, see [1, Fig. 2.5].

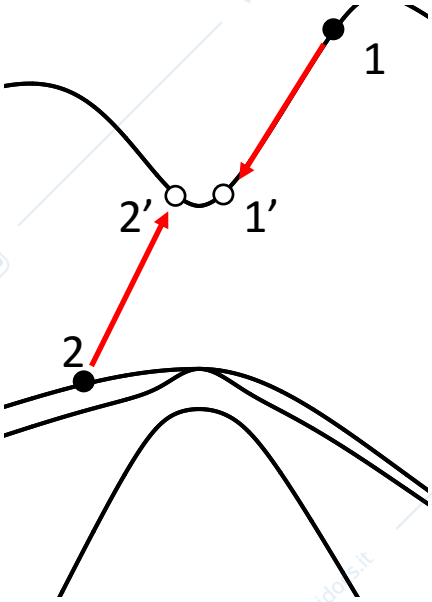
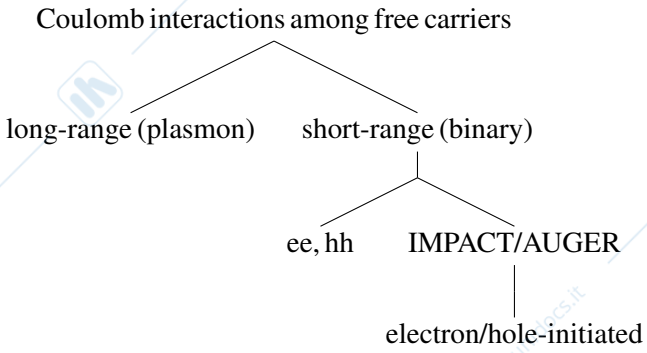


Figure 3.19.: Schematic view of an electron-initiated impact ionization process in GaAs. The initial electrons in the conduction and valence bands are in states 1 and 2 respectively; the final electrons in the conduction band after the transition are in states $1'$ and $2'$. Electrons are represented by closed circles, holes by open circles. Arrows indicate electron transitions.



3.3. Impact ionization (Auger generation)

In high-field conditions, electrons and holes gather enough energy from the electric field between two successive scattering events (e.g., collisions with phonons or impurities) to be able to interact with another electron (hole) and promote it to the conduction (valence) band. Each electron or hole is therefore able to generate, over a certain length, a number of electron-hole pairs, which undergo in turn the same process. The resulting chain can lead to a divergent current, i.e., to avalanche breakdown in the semiconductor. Avalanche breakdown occurs for electric fields of the order of the breakdown field ($\approx 500\,\mathrm{kV/cm}$ for GaAs), and is usually an undesirable effect, with a notable exception for avalanche photodetectors (APDs). A threshold is clearly involved in this process since the initiating particle must have an energy higher than the band gap to provide enough energy for the valence electron to cross the gap. Moreover, this threshold may be further increased by the momentum conservation condition or by the unavailability of energy conserving particle states. For these reasons it is expected that the impact ionization rates, particularly at threshold, are highly dependent on the band structure.

In the impact-ionization process, an electron in the conduction band interacts with an electron in the

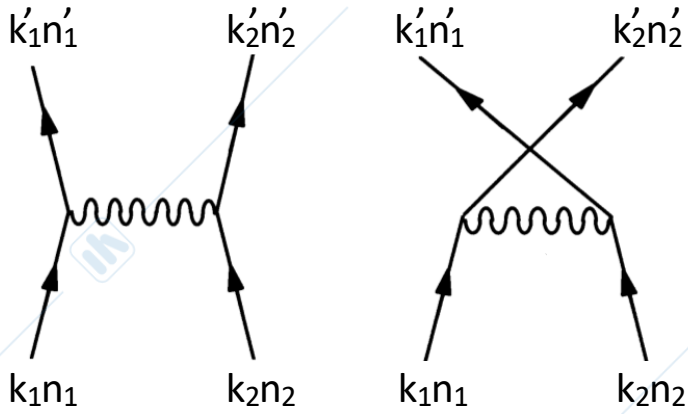


Figure 3.20.: Feynman's diagrams of the direct and exchange processes.

valence band via the screened Coulomb potential

$$V(\underline{r}_1 - \underline{r}_2) = \frac{e^2}{4\pi\varepsilon(\underline{q},\omega)} \frac{1}{|\underline{r}_1 - \underline{r}_2|} \tag{3.89}$$

If states 1 (a Bloch state $\psi_{\underline{k}_1}^{(n_1)}(\underline{r}_1)$ with band index n_1 and wavevector $\underline{k}_1$) and 2 (a Bloch state $\psi_{\underline{k}_2}^{(n_2)}(\underline{r}_2)$ with band index n_2 and wavevector $\underline{k}_2$) are the initial states of the electrons in the conduction and valence band before the transition, respectively, and states $1'$ and $2'$ are those of the electrons in the conduction band after the transition, the total ionization rate is given by [4, Chapter 15]

$$\begin{aligned} S_{\mathrm{ii}}(\underline{k}_1 n_1, \underline{k}_2 n_2; \underline{k}_{1'} n_{1'}, \underline{k}_{2'} n_{2'}) = \frac{2\pi}{\hbar} |M(\underline{k}_{1'} n_{1'} \underline{k}_{2'} n_{2'}; \underline{k}_1 n_1 \underline{k}_2 n_2)|^2 \\ \times \delta[E_{n_1}(\underline{k}_1) + E_{n_2}(\underline{k}_2) - E_{n_{1'}}(\underline{k}_{1'}) - E_{n_{2'}}(\underline{k}_{2'})]. \end{aligned} \tag{3.90}$$

Summing over all spin configurations, it can be shown that the matrix element M includes a direct term M_D and an exchange term M_E obtained by simply exchanging the wavevectors and the band indices of the final states [26, Chapter 3]

$$|M|^2 = \frac{1}{2} \left(\underbrace{|M_D|^2 + |M_E|^2}_{\text{unlike spin}} + \underbrace{|M_D - M_E|^2}_{\text{like spin}} \right), \tag{3.91}$$

with

$$M_D = \iint \mathrm{d}\underline{r}_1 \mathrm{d}\underline{r}_2 \; \psi_{1'}^*(\underline{r}_1) \psi_{2'}^*(\underline{r}_2) V(\underline{r}_1 - \underline{r}_2) \psi_1(\underline{r}_1) \psi_2(\underline{r}_2) \tag{3.92a}$$

$$M_E = \iint \mathrm{d}\underline{r}_1 \mathrm{d}\underline{r}_2 \; \psi_{2'}^*(\underline{r}_1) \psi_{1'}^*(\underline{r}_2) V(\underline{r}_1 - \underline{r}_2) \psi_1(\underline{r}_1) \psi_2(\underline{r}_2). \tag{3.92b}$$

Electrons 1 and 2 have different spin in the terms $|M_D|^2$ and $||M_E|^2$, while they have the same spin in $|M_D - M_E|^2$. Expressing the interacting potential in Fourier series

$$V(\underline{r}_1 - \underline{r}_2) = \frac{e^2}{4\pi\varepsilon(\underline{q},\omega)} \frac{1}{\Omega} \sum_{\underline{q}} \frac{4\pi}{q^2} e^{\mathrm{i}\underline{q}\cdot(\underline{r}_1 - \underline{r}_2)}, \tag{3.93}$$

and remembering the expression of the Bloch states

$$\psi_{\underline{k}}^{(n)}(\underline{r}) = \frac{1}{\sqrt{\Omega}} e^{\mathrm{i}\underline{k}\cdot\underline{r}} \sum_{\underline{G}} u_{\underline{G}\underline{k}}^{(n)} \, e^{\mathrm{i}\underline{G}\cdot\underline{r}}, \tag{3.94}$$

the direct matrix element M_D becomes

$$M_D = \sum_{\underline{G}_1\underline{G}_2\underline{G}_{1'}\underline{G}_{2'}} u^{(n_{1'})*}_{\underline{G}_{1'}\underline{k}_{1'}} u^{(n_{2'})*}_{\underline{G}_{2'}\underline{k}_{2'}} u^{(n_1)}_{\underline{G}_1\underline{k}_1} u^{(n_2)}_{\underline{G}_2\underline{k}_2} \frac{1}{\Omega}\sum_{\underline{q}} \frac{e^2}{\varepsilon(\underline{q},\omega)q^2} \tag{3.95}$$

$$\times \underbrace{\frac{1}{\Omega^2}\iint \mathrm{d}\underline{r}_1 \mathrm{d}\underline{r}_2\, e^{\mathrm{i}(\underline{k}_1+\underline{G}_1-\underline{k}_{1'}-\underline{G}_{1'}+\underline{q})\cdot\underline{r}_1} e^{\mathrm{i}(\underline{k}_2+\underline{G}_2-\underline{k}_{2'}-\underline{G}_{2'}-\underline{q})\cdot\underline{r}_2}}_{\delta(\underline{k}_1+\underline{G}_1-\underline{k}_{1'}-\underline{G}_{1'}+\underline{q})\,\delta(\underline{k}_2+\underline{G}_2-\underline{k}_{2'}-\underline{G}_{2'}-\underline{q})}. \tag{3.96}$$

Integration over $\underline{r}_1$ and $\underline{r}_2$ gives

$$M_D = \frac{1}{\Omega}\sum_{\underline{G}_1\underline{G}_2\underline{G}_{1'}\underline{G}_{2'}} u^{(n_{1'})*}_{\underline{G}_{1'}\underline{k}_{1'}} u^{(n_{2'})*}_{\underline{G}_{2'}\underline{k}_{2'}} u^{(n_1)}_{\underline{G}_1\underline{k}_1} u^{(n_2)}_{\underline{G}_2\underline{k}_2} \frac{e^2}{\varepsilon(\underline{q}_D,\omega_D)q_D^2} \tag{3.97}$$

$$\times\, \delta(-\underline{k}_{1'}-\underline{G}_{1'}+\underline{k}_1+\underline{G}_1-\underline{k}_{2'}-\underline{G}_{2'}+\underline{k}_2+\underline{G}_2) \tag{3.98}$$

where

$$\hbar\omega_D = E_{n_1}(\underline{k}_1) - E_{n_{1'}}(\underline{k}_{1'}) \tag{3.99}$$

$$\underline{q}_D = \underline{k}_{1'}+\underline{G}_{1'}-\underline{k}_1-\underline{G}_1 = \underline{k}_{2'}+\underline{G}_{2'}-\underline{k}_2-\underline{G}_2 \tag{3.100}$$

are the direct energy and momentum transfer, respectively. Simular expressions hold for the exchange matrix element M_E. Integrating over the final states of the initiating particle and over the initial and final states of the partner particle we obtain the scattering rate

$$W_{\mathrm{ii}}(\underline{k}_1 n_1) = \frac{2\pi}{\hbar}\sum_{\underline{k}_{1'}n_{1'}\underline{k}_{2'}n_{2'}\underline{k}_2 n_2} |M(\underline{k}_{1'}n_{1'}\underline{k}_{2'}n_{2'};\underline{k}_1 n_1 \underline{k}_2 n_2)|^2$$

$$\times\, \delta[E_{n_1}(\underline{k}_1)+E_{n_2}(\underline{k}_2)-E_{n_{1'}}(\underline{k}_{1'})-E_{n_{2'}}(\underline{k}_{2'})]. \tag{3.101}$$

The major difficulty in the evaluation of impact ionization matrix elements M_D and M_E lies in choosing an accurate yet manageable expression for the dielectric function. In general, any external perturbation causes not only the scattering process we are interested in, but also a *many-body* rearrangement of all electrons, valence and conduction, as well as of the ions (especially if the crystal is ionic). This rearrangement modifies the scattering potential, so that the scattering event depends on these many-body effects as well. Since the energy exchanged in impact ionization processes are typically much larger than phonon energies, we consider here only the dielectric response of the valence electrons. Within the random phase approximation (RPA), the valence band contribution to the wavevector-dependent dielectric function takes the usual Lindhard expression [27, Chapter 4]

$$\varepsilon(\underline{q},\omega) = \varepsilon_0 + \frac{e^2}{\Omega q^2}\sum_{\underline{k},c,v} |\langle \underline{k},c|\underline{k}+\underline{q},v\rangle|^2$$

$$\times\{[E_c(\underline{k})-E_v(\underline{k}+\underline{q})-\hbar\omega-\mathrm{i}\eta]^{-1}+[E_c(\underline{k})-E_v(\underline{k}+\underline{q})+\hbar\omega+\mathrm{i}\eta]^{-1}\}, \tag{3.102}$$

where ε_0 is the permittivity of vacuum, $\underline{k}$ is summed over the first Brillouin zone, v labels the topmost (occupied) valence bands, c labels the lowest-lying (empty) conduction bands, and η is a positive and infinitesimally small value, which is eliminated at the end of the calculations by taking the limit $\eta \to 0$ (i.e, we are computing the principal value of the integral).

The multidimensional integration in (3.101) is performed by a Monte Carlo technique. The integral is typically slow to converge due to the fact that only a very small fraction of randomly sampled final state pairs satisfy the approximate energy and momentum conservation condition, especially when the impacting vector $\underline{k}_1$ lies near the impact threshold. Fig. 3.21 presents the electron-initiated impact ionization transition rate for wurtzite GaN in 3402 k-points of the irreducible wedge. A significant spread in the value of the scattering rate at a fixed energy is evident, indicating that an energy-dependent averaging of the scattering rate (solid line) would not properly reflect the strong $\underline{k}$-dependence of the scattering

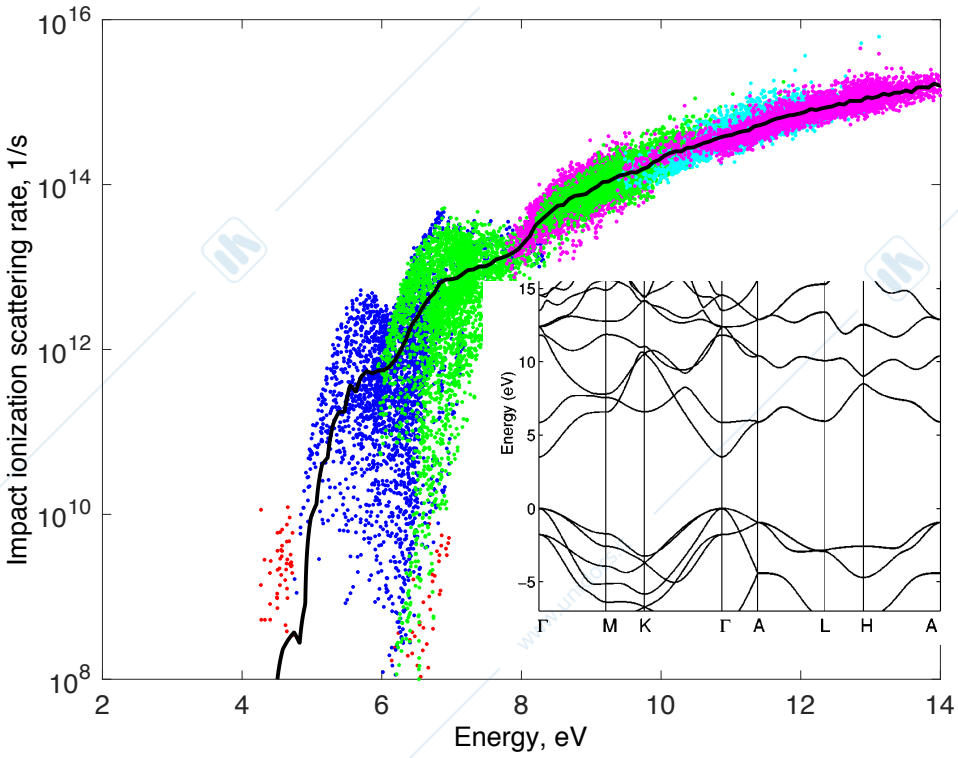


Figure 3.21.: Calculated impact ionization transition rate for electrons in wurtzite GaN. Each symbol represents the rate for a $\underline{k}$-point corresponding to a given energy and band. Different colors are associated to different conduction bands. The solid line is the energy averaged rate.

mechanism. The anisotropy of the ionization probability may be traced back to the strong restrictions imposed by energy and momentum conservation in the evaluation of the matrix element for a realistic band structure.

According to the principle of detailed balance, at equilibrium, the number of carriers that appear per unit time due to a generation process must be equal to the number of carriers that disappear due to the inverse recombination process. The inverse process of impact ionization si direct (i.e., phonon-less) Auger recombination[7]. The matrix element is therefore exactly the same[8]. Auger recombination involves either two electrons and a hole (electron-electron-hole, *eeh*) or two holes and an electron (hole-hole-electron, *hhe*). The *eeh* (*hhe*) recombination energy excites a second electron (hole) to a higher state, which can be either in the same or in a higher band. The Auger rate can be computed as [33]

$$R = 2\frac{2\pi}{\hbar}\sum_{\underline{k}_1 n_1 \underline{k}_{1'} n_{1'} \underline{k}_2 n_2 \underline{k}_{2'} n_{2'}} |M(\underline{k}_{1'} n_{1'} \underline{k}_{2'} n_{2'}; \underline{k}_1 n_1 \underline{k}_2 n_2)|^2 \delta[E_{n_1}(\underline{k}_1) + E_{n_2}(\underline{k}_2) - E_{n_{1'}}(\underline{k}_{1'}) - E_{n_{2'}}(\underline{k}_{2'})]$$
$$\times f(E_{n_1}(\underline{k}_1)) f(E_{n_2}(\underline{k}_2))[1 - f(E_{n_{1'}}(\underline{k}_{1'}))][1 - f(E_{n_{2'}}(\underline{k}_{2'}))], \quad (3.103)$$

[7] When, in 1922, Lise Meitner described radiationless transitions to explain the emission of electrons with signature energies as a corollary of an article on nuclear physics [31], a remarkable discovery went unnoticed. The effect is known after Pierre Auger, who independently discovered it in 1923 [32]. Auger recombination, which should have been named at least *Meitner-Auger* recombination, is now at the heart of the debate concerning "droop", the decline of the internal quantum efficiency in GaN-based LEDs at high injection currents.

[8] Since impact ionization (II) and Auger (AR) are inverse microscopic processes related by the detailed balance principle, the Auger coefficient provides qualitative information about the overall strength of the II process. A direct connection between II and Auger coefficients (the average number of electron-hole pairs generated per unit length) cannot be made, because both quantities depend on the carriers distribution: Auger coefficients, or equivalently the Auger lifetimes, are usually computed assuming that the electron and hole populations are described by quasi-Fermi levels, while II coefficients are usually evaluated in high field conditions.

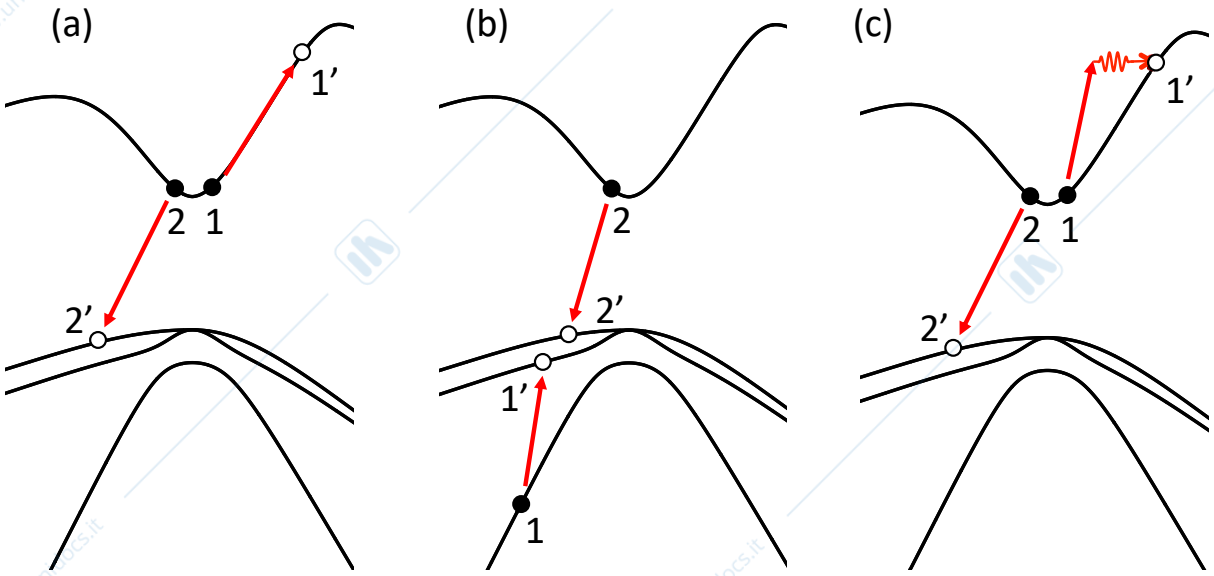


Figure 3.22.: Band to band Auger transitions in GaAs. (a) *eeh* process; (b) *hhe* process. Electrons are represented by closed circles, holes by open circles. Arrows indicate electron transitions. Process (a) is the reverse of the process in Fig. 3.19. In thermal equilibrium the recombination and generation rates balance, while, away from equilibrium, the net recombination rate is their difference. (c) Phonon-assisted *eeh* Auger process. The phonon relaxes energy and momentum conservation rules and enables Auger transitions to a wider range of final states in the Brillouin zone.

where $f(E)$ is the probability that an electron is occupying the state with energy E. (For *hhe* processes $f(E)$ is the occupation probability for a hole.) The integrals over $\underline{k}_1$, $\underline{k}_2$, $\underline{k}_3$, $\underline{k}_4$ span 12 dimensions; the momentum-conserving δ function included in the matrix element M can be used to eliminate the integration over $\underline{k}_2$, leaving a nine-dimensional integral. For non-degenerate semiconductors, for which Boltzmann statistics applies, the *eeh* Auger rate is proportional to n^2p, as the occupation probability for electrons (holes) is proportional to their concentration n (p). Likewise the *hhe* process is proportional to p^2n. Accordingly, we can define the Auger coefficients C_n, C_p

$$R_{eeh} = C_n n^2 p \tag{3.104}$$

$$R_{hhe} = C_p p^2 n. \tag{3.105}$$

The strong temperature dependence of direct Auger process arises from carrier distribution statistics and momentum conservation, which leads to the well known Auger threshold condition. For the same reason, the direct Auger rate decreases rapidly with increasing band gap. Phonon-assisted processes are similar to the corresponding direct transitions, but one of the involved carriers also scatters with a phonon, see Fig. 3.22c. Phonon-dressed recombination exhibits a weaker temperature dependence, since momentum conservation is lifted at the expense of a large momentum transferred to the phonon. In wide band gap semiconductors, due to the large threshold of Auger transitions and to the exponential behavior of the statistical factors accounting for the occupation probability of initial and final states, indirect recombination processes dominate over direct ones. They also remain effective at low temperatures, since a thermal phonon population is not required by phonon-emission-assisted transitions. At energies lower than the threshold, Auger recombination is not possible without the contribution of other scattering mechanisms. In addition to phonon-assisted Auger processes, different transitions have been considered: alloy-, trap-assisted Auger processes, combinations of Auger and SRH processes, and more, see Fig. 3.23.

As already mentioned in footnote 7, Auger recombination is one of the most debated mechanisms invoked as a possible origin of the efficiency droop affecting nitride-based light-emitting diodes (LEDs) at high injection currents, see Fig. 3.24. A great deal of effort has been put in the analysis of the efficiency

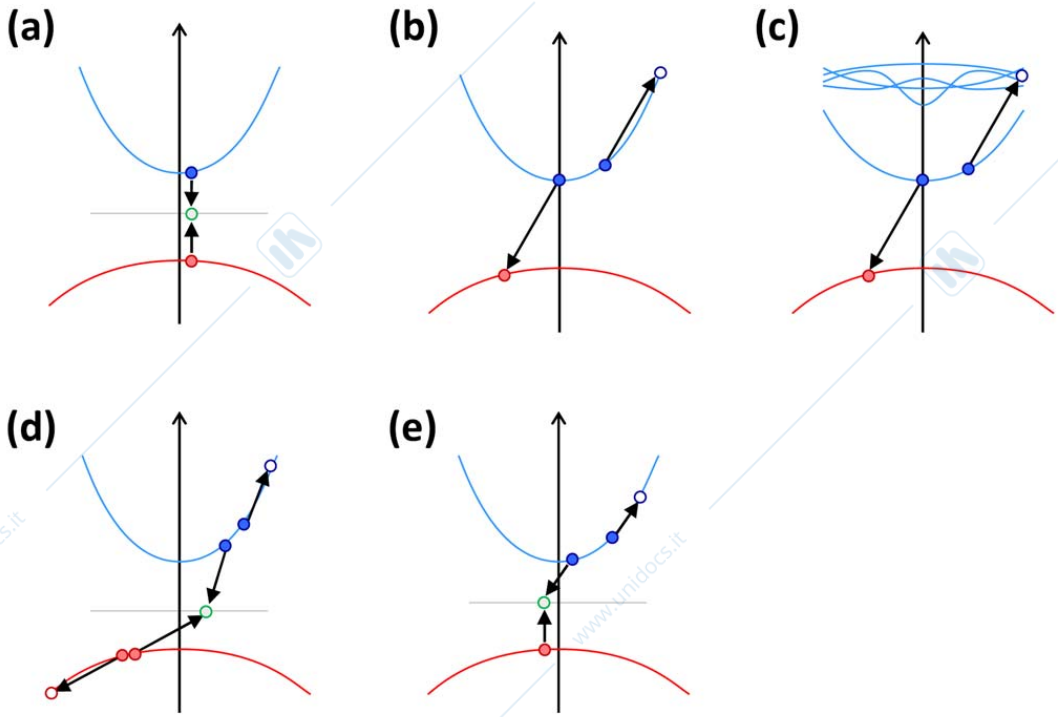


Figure 3.23.: Simplified diagrams for non-radiative scattering processes. (a) SRH recombination consisting of two multiphonon-recombination events. (b) Direct *eeh* Auger scattering. (c) Alloy-assisted Auger scattering, where composition fluctuations create multiple subbands. (d) Two trap-assisted Auger processes, where both an electron and a hole are captured by a trap. (e) Combination of a trap-assisted process and an SRH process. Reproduced from [28].

droop, the decrease in LED internal quantum efficiency (IQE) induced by high current density. Only a fraction of the electrons crossing the junction are effectively captured by the quantum well. The remaining part can transfer to the *p*-side either via leakage through/over the EBL or through defect-related paths, depending on the current level and on device structure. At low current densities, Shockley-Read-Hall recombination plays the strongest role: the presence of defects within the active region can promote defect-assisted non-radiative recombination, thus leading to a decrease in the internal quantum efficiency. In the quantum well, the energy levels are quantized, and electrons mostly reside in the ground state, but a part of the carriers may have enough energy to escape from the active region. The parameter that controls the escape is the conduction/valence band discontinuity ΔE_C and ΔE_V, see Fig. 3.25. Auger recombination may also contribute to carrier overflow. In fact, Auger-generated electrons have a significant probability of leaking above the EBL, and reaching the extra quantum well located on the *p*-side. An IQE decrease is observed also for increasing temperature, a phenomenon usually referred to as *thermal droop*. For commercial LEDs, the IQE decrease related to thermal droop can be comparable to that of efficiency droop. The analysis of the underlying processes is not straightforward, since both recombination-related processes (e.g. changes in the recombination coefficients) and transport-related phenomena (including carrier delocalization and escape) can contribute. In the Quixotic search for experimental evidence of Auger losses, two notable experiments should be mentioned, see Figs. 3.26 and 3.27.

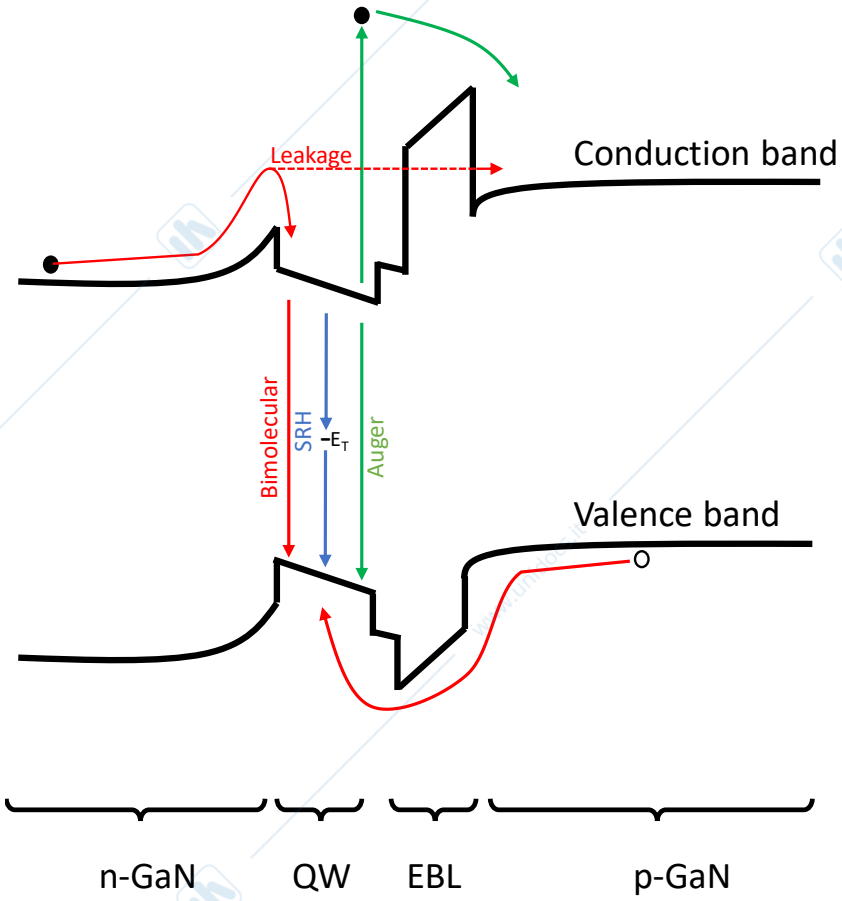


Figure 3.24.: Schematic band diagram of a single quantum well GaN LED, indicating the dominant loss mechanisms that limit the internal quantum efficiency: SRH recombination, Auger recombination, carrier leakage.

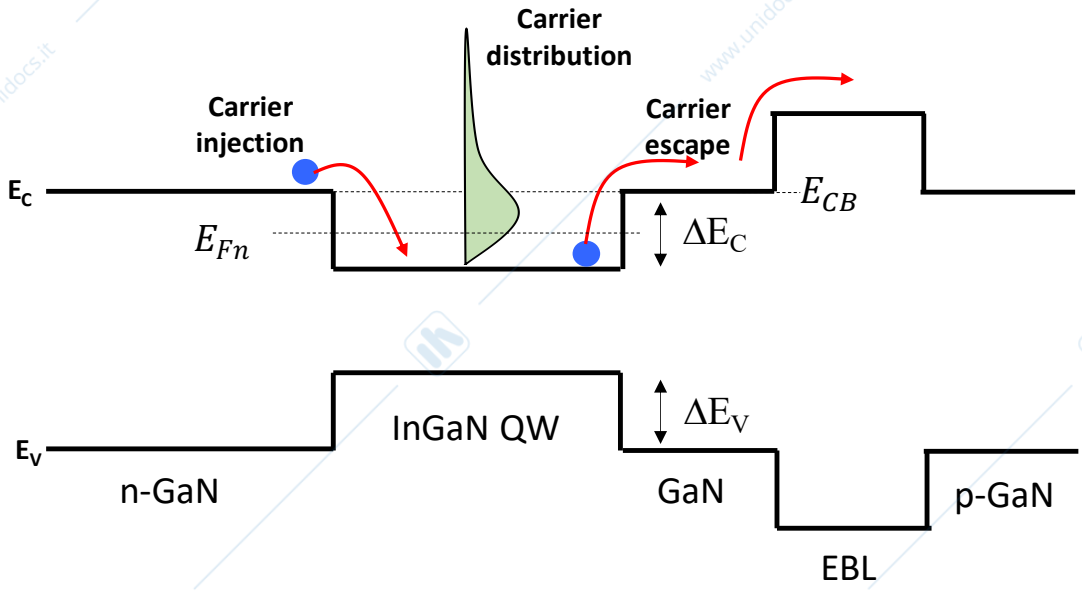


Figure 3.25.: Schematic band diagram of a single quantum well GaN LED, focusing on the carrier injection and escape processes.

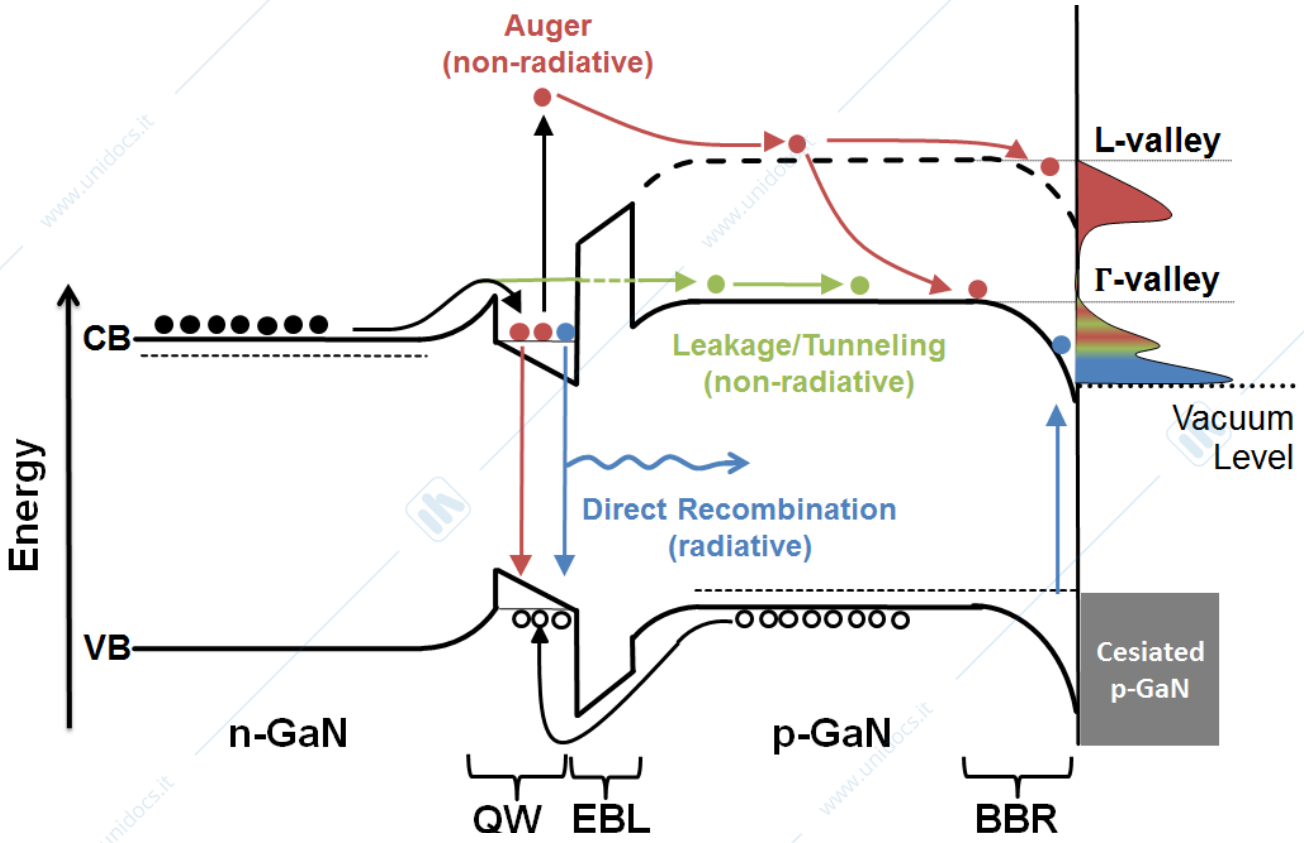


Figure 3.26.: A spectroscopic study of the electrons emitted from the *p*-cap of a forward-biased GaN-based LED, revealed a high-energy peak, attributed to hot electrons generated by Auger recombination in the active region and supposedly thermalized at the bottom of a satellite valley in GaN [29].

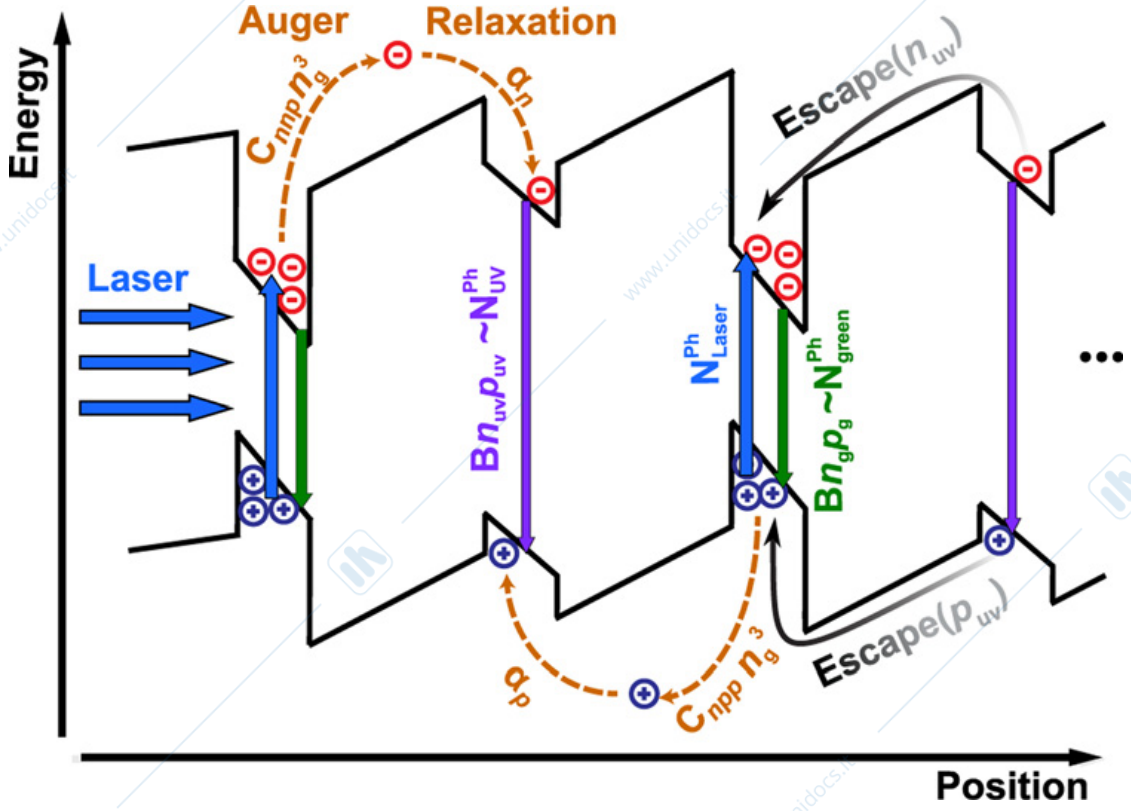


Figure 3.27.: A test structure composed of alternating ultraviolet (UV) and green quantum wells was optically pumped into droop regime using a blue high-power laser source. By optically pumping solely the green QWs using a blue emitting high power laser diode, carrier densities similar to electrical LED operation were achieved, circumventing possible leakage and injection effects. Luminescence from the UV QWs was observed for excitation where the emission from the green QWs showed significant droop, giving direct evidence for hot electrons and holes, promoted to higher energies by Auger transitions, escaping the green QWs and subsequently captured by the UV QWs [30].

4. Solving the Boltzmann equation

4.1. Linear transport: relaxation time approximation

The scattering rate for an electron in a well-defined state is virtually never directly observable. In almost all situations in practice, what is within the reach of experiments are phenomenological quantities such as mobility and diffusion coefficients, which are only indirectly related to the scattering rates. These quantities are properties of a population of electrons rather than of a single electron.

When an electric field is applied to a semiconductor material, charge carriers are accelerated along the force direction, while, at the same time, collisions tend to restore the equilibrium momentum distribution. A stationary situation is set up where the distribution function in k-space is shifted along the force direction, resulting in a current proportional to the displacement along the electric field direction, see Fig. 4.1. If the applied fields are sufficiently weak, the system is only slightly out of equilibrium, and the BTE can be linearized with respect to the external fields by means of linear response theory. Let's write the carrier distribution function as the sum of its equilibrium Fermi distribution $f_0(\underline{k})$, plus the first-order (w/r/t the electric field $\underline{\mathscr{E}}$) term $\underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}$

$$f(\underline{k}) = f_0(\underline{k}) + \underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}. \tag{4.1}$$

In stationary and homogeneous conditions, i.e., when the distribution function and the scattering mechanisms are independent of position and time, the BTE (1.1) simplifies to

$$\cancel{\frac{\partial f}{\partial t}} + \underbrace{\cancel{\underline{v} \cdot \nabla_{\underline{r}} f}}_{\text{diffusion}} + \underbrace{\frac{1}{\hbar} \underline{F} \cdot \nabla_{\underline{k}} f}_{\text{drift}} = \left.\frac{\partial f}{\partial t}\right|_{\text{coll.}} \tag{4.2}$$

Insertion of the ansatz (4.1) leads to

$$\begin{aligned} -\frac{e\underline{\mathscr{E}}}{\hbar} \cdot \nabla_{\underline{k}} [f_0(\underline{k}) + \underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}] = \sum_{\underline{k}'} \{ & [f_0(\underline{k}') + \underline{f}_1(\underline{k}') \cdot \underline{\mathscr{E}}] \, S(\underline{k}', \underline{k}) \, [1 - f_0(\underline{k}) - \underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}] \\ & - [f_0(\underline{k}) + \underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}] \, S(\underline{k}, \underline{k}') \, [1 - f_0(\underline{k}') - \underline{f}_1(\underline{k}') \cdot \underline{\mathscr{E}}] \, \}. \end{aligned} \tag{4.3}$$

In equilibrium conditions, without external forces ($\underline{\mathscr{E}} = 0$), the BTE simplifies to

$$\sum_{\underline{k}'} \{ f_0(\underline{k}') \, S(\underline{k}', \underline{k}) \, [1 - f_0(\underline{k})] - f_0(\underline{k}) \, S(\underline{k}, \underline{k}') \, [1 - f_0(\underline{k}')] \, \} = 0. \tag{4.4}$$

So, zero-order terms cancel for the equilibrium condition. Neglecting second-order terms, we are left with

$$\begin{aligned} -\frac{e\underline{\mathscr{E}}}{\hbar} \cdot \nabla_{\underline{k}} f_0(\underline{k}) = \sum_{\underline{k}'} \{ & [\underline{f}_1(\underline{k}') \cdot \underline{\mathscr{E}}] \, S(\underline{k}', \underline{k}) \, [1 - f_0(\underline{k})] - f_0(\underline{k}') \, S(\underline{k}', \underline{k}) \, \underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}} \\ & - [\underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}] \, S(\underline{k}, \underline{k}') \, [1 - f_0(\underline{k}')] + f_0(\underline{k}) \, S(\underline{k}, \underline{k}') \, [\underline{f}_1(\underline{k}') \cdot \underline{\mathscr{E}}] \, \} \end{aligned} \tag{4.5}$$

The detailed-balance principle states that the integral in (4.4) is zero because the integrand itself is zero, i.e., each transition occurs, at equilibrium, with the same frequency as the opposite transition:

$$f_0(\underline{k}') \, S(\underline{k}', \underline{k}) \, [1 - f_0(\underline{k})] = f_0(\underline{k}) \, S(\underline{k}, \underline{k}') \, [1 - f_0(\underline{k}')]. \tag{4.6}$$

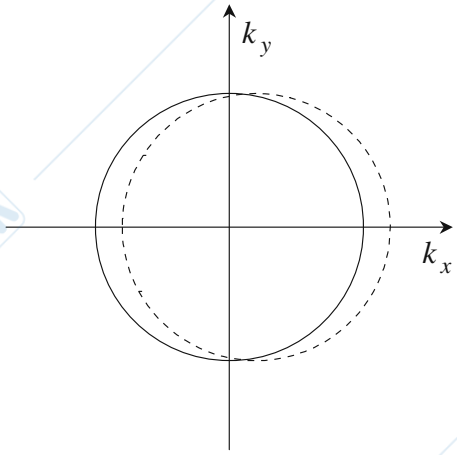


Figure 4.1.: Change in the distribution function under an applied electric field. In the absence of an electric field, the distribution function of the thermal equilibrium state (solid line) is isotropic in momentum space and thus the current in any direction has the same magnitude, resulting in mutual cancellation and hence zero current. In the presence of an applied electric field, the distribution function is given by $f(\underline{k}) = f_0(\underline{k}) + \underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}$ (dashed line) and is displaced by $\underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}}$ in the electric field direction, resulting in a current proportional to the displacement along the electric field direction. Reproduced from [34].

With the explicit form of the Fermi distribution

$$f_0(\underline{k}) = \frac{1}{e^{\beta(E_{\underline{k}}-\mu)}+1}, \tag{4.7}$$

the ratio between a transition probability and the probability of the opposite transition is given by the Boltzmann factor of their energy difference

$$\frac{S(\underline{k},\underline{k}')}{S(\underline{k}',\underline{k})} = \frac{f_0(\underline{k}')[1-f_0(\underline{k})]}{f_0(\underline{k})[1-f_0(\underline{k}')]} = e^{-\beta(E_{\underline{k}'}-E_{\underline{k}})}, \tag{4.8}$$

where $\beta = 1/k_B T$. By using the detailed-balance expression (4.8), the linearized BTE (4.5) becomes

$$\begin{aligned} -\frac{e\underline{\mathscr{E}}}{\hbar} \cdot \nabla_{\underline{k}} f_0(\underline{k}) = \sum_{\underline{k}'} \{ & \underline{f}_1(\underline{k}') \cdot \underline{\mathscr{E}} S(\underline{k},\underline{k}') \left[(1-f_0(\underline{k}))e^{\beta(E_{\underline{k}'}-E_{\underline{k}})} + f_0(\underline{k})\right] \\ & - \underline{f}_1(\underline{k}) \cdot \underline{\mathscr{E}} S(\underline{k},\underline{k}') \left[(1-f_0(\underline{k}')) + f_0(\underline{k}')e^{\beta(E_{\underline{k}'}-E_{\underline{k}})}\right] \} \end{aligned} \tag{4.9}$$

Analytical solutions are possible if collision integral in the the r.h.s. of (4.9) can be written in a form known as relaxation time approximation

$$\left.\frac{\partial f}{\partial t}\right|_{\text{coll.}} = -\frac{f(\underline{k}) - f_0(\underline{k})}{\tau(\underline{k})}, \tag{4.10}$$

which describes in a phenomenological way the relaxation of the system towards the equilibrium distribution. A relaxation time $\tau(\underline{k})$ can be rigorously defined in two cases: (1) velocity randomizing (isotropic) scattering processes and (2) elastic scattering processes

1. Collisions are velocity randomizing when the final state of the collision is equally probable with respect to the final state with opposite velocity (deformation potential scattering belongs to this category). Then, time reversal symmetry (opposite velocities correspond to opposite wavevectors)

implies $S(\underline{k},\underline{k}') = S(\underline{k},-\underline{k}')$, and $\underline{f}_1(\underline{k}') = -\underline{f}_1(-\underline{k}')$, since $\underline{f}_1$ is proportional to $\underline{\mathscr{E}}$ (opposite electric fields correspond to opposite wavevectors). So, the integral in the first line of (4.9)

$$-\frac{e\underline{\mathscr{E}}}{\hbar}\cdot\nabla_{\underline{k}} f_0(\underline{k}) = \sum_{\underline{k}'}\{\underbrace{\underline{f}_1(\underline{k}')\cdot\underline{\mathscr{E}}\,S(\underline{k},\underline{k}')\,[(1-f_0(\underline{k}))e^{\beta(E_{\underline{k}'}-E_{\underline{k}})}+f_0(\underline{k})]}_{=0}$$
$$-\underline{f}_1(\underline{k})\cdot\underline{\mathscr{E}}\,S(\underline{k},\underline{k}')\,[(1-f_0(\underline{k}'))+f_0(\underline{k}')e^{\beta(E_{\underline{k}'}-E_{\underline{k}})}]\,\} \tag{4.11}$$

is zero since the integrand is an odd function of $\underline{k}'$, and we can write

$$-\frac{e\underline{\mathscr{E}}}{\hbar}\cdot\nabla_{\underline{k}} f_0(\underline{k}) = -\frac{\underline{f}_1(\underline{k})\cdot\underline{\mathscr{E}}}{\tau(\underline{k})}, \tag{4.12}$$

where $\tau(\underline{k})$ is a wavevector-dependent momentum relaxation time[1] given by

$$\frac{1}{\tau(\underline{k})} = \sum_{\underline{k}'}\left\{S(\underline{k},\underline{k}')[(1-f_0(\underline{k}'))+f_0(\underline{k}')e^{\beta(E_{\underline{k}'}-E_{\underline{k}})}]\right\}. \tag{4.13}$$

For nondegenerate statistics, when $f_0(\underline{k}') \ll 1$, the above reduces to

$$\frac{1}{\tau(\underline{k})} = \sum_{\underline{k}'} S(\underline{k},\underline{k}') \equiv W(E_{\underline{k}}), \tag{4.14}$$

which has a simple physical interpretation: when the scattering probability is the same for opposite velocities of the state after scattering, at each collision the electron momentum is completely dissipated, in average, and the momentum relaxation time coincides with the inverse scattering rate.

2. The collision is elastic when initial and final energies are the same (impurity scattering belongs to this category). Following a similar procedure, it can be shown that (θ is the scattering angle between $\underline{k}$ and $\underline{k}'$, see Fig. 3.3)

$$\frac{1}{\tau(\underline{k})} = \sum_{\underline{k}'} S(\underline{k},\underline{k}')[1-\cos\theta], \tag{4.15}$$

which also has a direct physical meaning: if the scattering rate is not velocity randomizing, and the modulus of the momentum remains unchanged, only a fraction $(1-\cos\theta)$ of the momentum is lost in the collision, and the momentum relaxation time is the average of the collision rate weighted by this fraction.

At this point, we are in a condition to evaluate the mobilities due to the various scattering mechanisms. Consider the l.h.s. of (4.12). Taking into account that the equilibrium distribution $f_0(\underline{k}) = f_0(E_{\underline{k}})$ depends on $\underline{k}$ through the energy E, the chain derivation rule delivers

$$\frac{1}{\hbar}\nabla_{\underline{k}} f_0(\underline{k}) = \frac{1}{\hbar}\frac{\partial f_0}{\partial E}\nabla_{\underline{k}} E(\underline{k}) = \frac{\partial f_0}{\partial E}\,\underline{v}(\underline{k}). \tag{4.16}$$

Substituting in (4.12), we find

$$\underline{f}_1(\underline{k}) = e\frac{\partial f_0}{\partial E}\,\tau(\underline{k})\,\underline{v}(\underline{k}) \tag{4.17}$$

[1] The reason of the name "relaxation time" is due to the fact that if we drag the distribution function f out of equilibrium, and then we remove any applied force at $t=0$, f "relaxes" to its equilibrium value under the action of the collisions according to $f_1(\underline{k},t) = f_1(\underline{k},0)e^{-t/\tau}$. In linear response regime, τ relaxes only momentum, or velocity, because the energy distribution is still the equilibrium one. Far from equilibrium an energy relaxation time can also be defined. Momentum and energy relaxation times are important parameters needed in hydrodynamic calculations.

In the nondegenerate limit $E_{\underline{k}} - \mu \gg k_B T$, the Fermi-Dirac distribution f_{FD} can be approximated by the Maxwell-Boltzmann distribution f_M (the exponential in the denominator becomes dominant with respect to unity)

$$f_{\mathrm{FD}}(\underline{k}) = \frac{1}{e^{\beta(E_{\underline{k}}-\mu)}+1} \approx f_{\mathrm{M}}(\underline{k}) = e^{-\beta(E_{\underline{k}}-\mu)} \tag{4.18}$$

whose derivative with respect to energy is

$$\partial f_{\mathrm{M}}/\partial E = -\beta f_{\mathrm{M}}, \tag{4.19}$$

which gives

$$f(\underline{k}) = f_0(\underline{k}) \underbrace{-e\beta f_0(\underline{k})\tau(\underline{k})\underline{v}(\underline{k})}_{f_{-1}(\underline{k})} \cdot \underline{\mathscr{E}}. \tag{4.20}$$

Inserting (4.20) in the expression of the average velocity over the distribution function (the *drift* velocity)

$$\underline{v}_{\mathrm{d}} = \langle \underline{v}(\underline{k}) \rangle_{\underline{k}} = \frac{\int \mathrm{d}\underline{k}\, f(\underline{k}) \underline{v}(\underline{k})}{\int \mathrm{d}\underline{k}\, f(\underline{k})}, \tag{4.21}$$

one finds for a parabolic band dispersion [2, Chapter 11]

$$\underline{v}_{\mathrm{d}} = -\mu_0 \underline{\mathscr{E}} = -\frac{e\underline{\mathscr{E}}}{m^*} \frac{\langle E\tau \rangle_{\underline{k}}}{\langle E \rangle_{\underline{k}}}, \tag{4.22}$$

which gives the *low-field* mobility

$$\mu_0 = \frac{e}{m^*} \frac{\langle E\tau \rangle_{\underline{k}}}{\langle E \rangle_{\underline{k}}}. \tag{4.23}$$

Multiplication by the density of states $N(E_{\underline{k}})$ converts the integrals over momentum space in (4.23) to integrals over energy space

$$\mu_0 = \frac{e}{m^*} \frac{2}{3k_B T} \frac{\int \mathrm{d}E\, f_{\mathrm{M}}(E)\, E\, \tau(E) N(E)}{\int \mathrm{d}E\, f_{\mathrm{M}}(E) N(E)} \tag{4.24}$$

which can be evaluated numerically by means of the Matlab routine *trapz*. The idea is to consider the different scattering mechanisms separately and then combine the corresponding mobilities by means of Matthiessen's rule. Acoustic scattering may be considered both velocity randomizing (isotropic) and elastic, see (3.45) so the momentum relaxation time is just the inverse scattering time

$$\frac{1}{\tau_{\mathrm{aco}}(E_{\underline{k}})} = W_{\mathrm{aco}}(E_{\underline{k}}). \tag{4.25}$$

Impurity scattering is not velocity randomizing as the scattering probability (3.87c) depends on the scattering angle θ, but it is an elastic process, so the corresponding relaxation time can be computed according to (4.15) and (3.87c)

$$\frac{1}{\tau_{\mathrm{imp}}^{(\mathrm{BH})}(E_{\underline{k}})} = \frac{\pi n_I Z^2 e^4 N(E_{\underline{k}})}{\hbar \varepsilon_s^2} \int_{-1}^{1} \frac{[1-\cos\theta]\mathrm{d}(\cos\theta)}{[2k^2(1-\cos\theta)+q_0^2]^2}, \tag{4.26}$$

which can be integrated analytically with the Symbolic Matlab Toolbox (assume complete ionization of the dopants)

$$\frac{1}{\tau_{\mathrm{imp}}^{(\mathrm{BH})}(E_{\underline{k}})} = \frac{\pi n_I Z^2 e^4 N(E_{\underline{k}})}{\hbar \varepsilon_s^2} \frac{1}{4k^4} \left[\log\left(1 + \frac{4k^2}{q_0^2}\right) - 1 \right]. \tag{4.27}$$

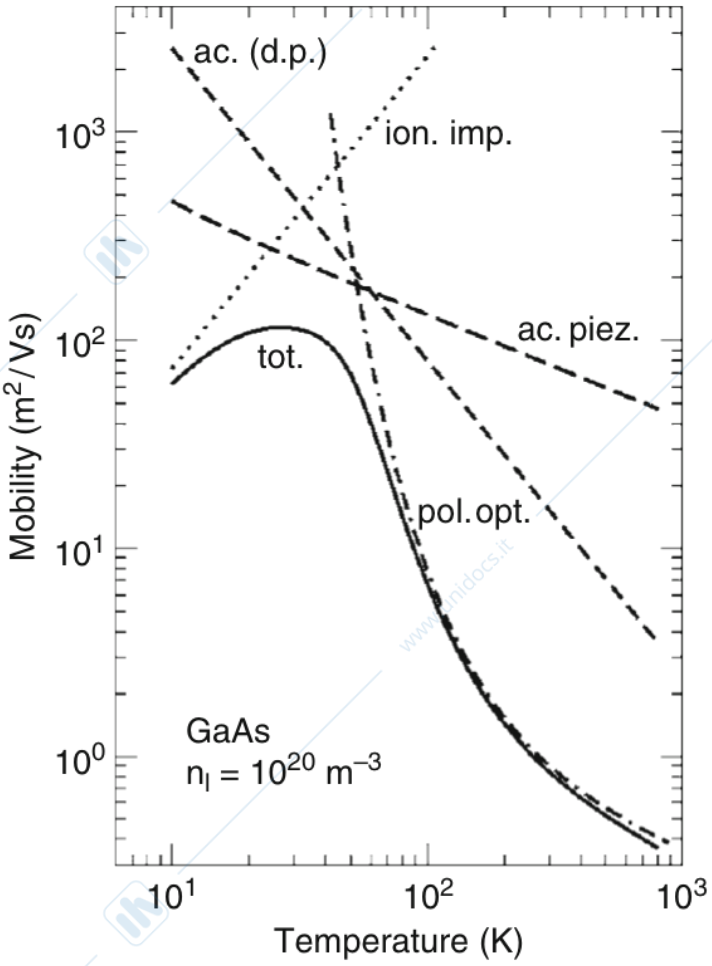


Figure 4.2.: Electron mobilities for different scattering mechanisms in GaAs, combined with Matthiessen rule. At room temperature, the mobility is controlled by polar optical phonons. At low temperatures, piezoelectric acoustic phonons are more effective. Ionized impurities may become important at low temperatures (assuming complete ionization of the dopands).

Thus all the processes considered here save for polar optical scattering and impact ionization,[2] admit a momentum relaxation time, and simple expressions of the corresponding mobilities are available. Polar optical scattering is neither elastic nor velocity randomizing, since the electrostatic nature of the interaction favors forward scattering. Strictly speaking a momentum relaxation time does not exist for polar optical scattering, although an approximate expression has been proposed [34, Chapter 6]

$$\frac{1}{\tau_{\text{pop}}(E_{\underline{k}})} = \frac{e^2\omega_{\text{LO}}}{8\pi\varepsilon_p}\frac{k}{E_{\underline{k}}}\left[(N_q+1)\left\{\sqrt{\frac{E_{\underline{k}}-\hbar\omega_{\text{LO}}}{E_{\underline{k}}}}+\frac{\hbar\omega_{\text{LO}}}{E_{\underline{k}}}\sinh^{-1}\sqrt{\frac{E_{\underline{k}}-\hbar\omega_{\text{LO}}}{\hbar\omega_{\text{LO}}}}\right\}\right.$$
$$\left.+N_q\left\{\sqrt{\frac{E_{\underline{k}}+\hbar\omega_{\text{LO}}}{E_{\underline{k}}}}-\frac{\hbar\omega_{\text{LO}}}{E_{\underline{k}}}\sinh^{-1}\sqrt{\frac{E_{\underline{k}}}{\hbar\omega_{\text{LO}}}}\right\}\right], \tag{4.28}$$

where it is understood that the emission term proportional to (N_q+1) is zero if $E_{\underline{k}} < \hbar\omega_{\text{LO}}$.

The theory we have developed here does a good job of describing the low-field electron transport in common semiconductors. Hole transport is more complicated because of the warped constant energy surfaces of light and heavy hole bands.

[2] N.B. that impact ionization is not considered here because it does not occur in low-field conditions where linear response theory applies.

Plot the total low-field mobility in GaAs due to impurity and deformation potential acoustic scattering as a function of temperature according to Matthiessen's rule $\mu_0^{-1} = \mu_{\text{aco}}^{-1} + \mu_{\text{imp}}^{-1} + \mu_{\text{pop}}^{-1}$. Compare the results with Fig. 4.2 (reproduced from [2, Fig. 11.7]).

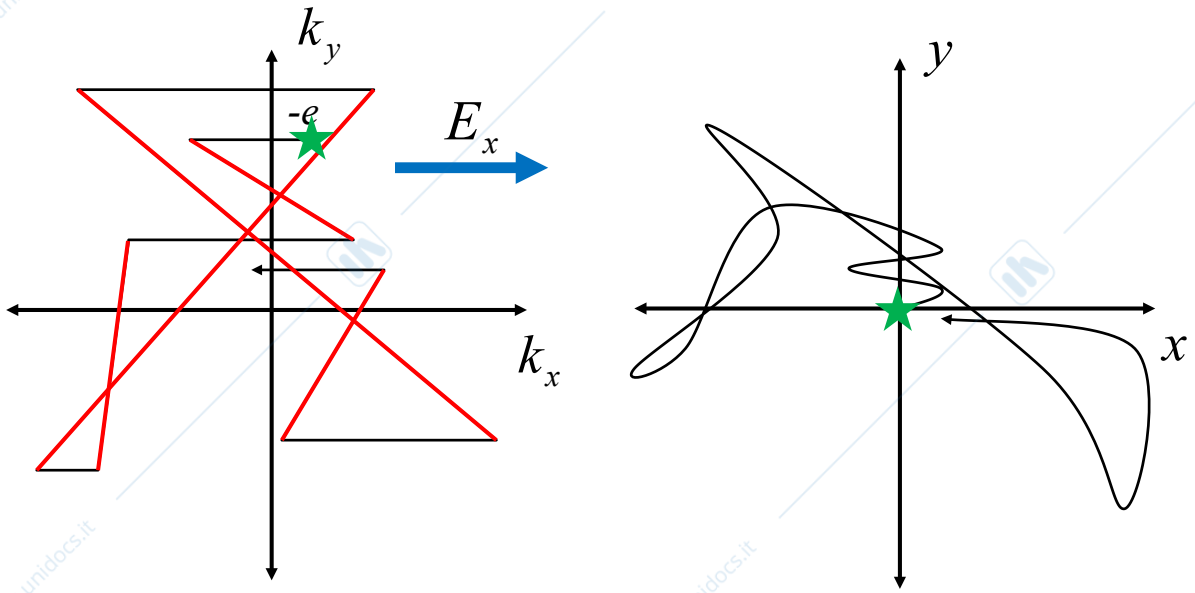


Figure 4.3.: Trajectories of an electron in real and momentum space. In the left panel, red lines represent scattering, black lines represent the propagation of the particle. N.B.: do not confuse instantaneous quantities with average quantities, e.g., instantaneous carrier velocities and the drift velocity. If you are interested in the history of the Monte Carlo method, have a look at Eric Grimson, John Guttag, and Ana Bell. 6.0002 Introduction to Computational Thinking and Data Science. Lecture 6: Monte Carlo Simulation. Fall 2016. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

4.2. Monte Carlo simulation of bulk properties of semiconductors

The motion of a particle can be conceived as the succession of drift and scattering processes, see Fig. 4.3. The electron wavevector $\underline{k}$ changes continuosly during a free flight because of the applied field. Because the scattering process is considered Markovian, the probability that a particle that scattered at time $t = 0$ has not yet suffered another collision at time t, i.e., the probability that the interval $(0,t)$ does not include any scattering event, is [35]

$$\mathscr{P}(t) = \exp\left[-\int_0^t W_T(E_{\underline{k}(t')})\mathrm{d}t'\right]. \tag{4.29}$$

Thus, the probability per unit time that a particle propagates under the effect of the electric field for a time interval t, at the end of which a scattering event occurs is

$$P(t) = W_T(E_{\underline{k}(t)}) \exp\left[-\int_0^t W_T(E_{\underline{k}(t')})\mathrm{d}t'\right], \tag{4.30}$$

where $W_T(E_{\underline{k}}) = \sum_{j=1}^{N} W_j(E_{\underline{k}})$ is the total scattering rate and the sum is over the N scattering mechanisms included in the simulation. The flight time τ may be selected stochastically by selecting a random number r uniformly distributed in the interval (0,1), and setting (Appendix C)

$$r = \int_0^\tau P(t)\mathrm{d}t. \tag{4.31}$$

Since this integral equation is typically troublesome to evaluate numerically, Rees proposed to add to the real scattering rates an artificial one $W_0(E_{\underline{k}})$ that makes the total scattering rate constant, i.e. independent

of $\underline{k}$

$$\Gamma = \sum_{j=0}^{N} W_j(E_{\underline{k}}). \tag{4.32}$$

Please note that the sum above now starts from zero to include the fictitious scattering mechanism (also known as self-scattering) that makes the total scattering rate equal to Γ. This constant value Γ should be chosen larger than the maximum value of the total scattering rate W_{T} in the energy range of interest (e.g., up to 1 eV), in order to avoid a negative value for $W_0(E_{\underline{k}})$.[3] Then, one simply has

$$P(t) = \Gamma \exp(-\Gamma t). \tag{4.33}$$

Having selected a pseudo-random[4] number r uniformly distributed in the open interval $(0,1)$ with the Matlab command *rand*, the flight time can then be determined with the inversion technique

$$r = \int_0^{\tau} P(t)\mathrm{d}t = \int_0^{\tau} \Gamma \exp(-\Gamma t)\mathrm{d}t, \tag{4.34}$$

which gives

$$\tau = -\ln(r)/\Gamma, \tag{4.35}$$

where we have used the fact that $(1-r)$ is stochastically equivalent to r. The *self-scattering technique* described above is very efficient in bulk Monte Carlo simulations that do not require a synchronous ensemble, but it is not the method of choice for more complicated simulations which involve e.g., carrier-carrier scattering, the self-consistent solution of the Poisson's equation, or the analysis of transients.

In this laboratory, we will use the *constant time technique*. The idea is to choose a constant flight time τ, and then decide a posteriori if a scattering event has occurred or not. If the free flight τ is short enough, $\underline{k}(t)$ can be considered constant during this time, and the probability (4.29) that no scattering occurs in the drift interval $(0,\tau)$ may be approximated as

$$\mathscr{P}(\tau) = \exp\left[-\int_0^{\tau} W_T(E_{\underline{k}(t')})\mathrm{d}t'\right] \approx \exp\left[-\tau W_T(E_{\underline{k}(\tau)})\right]. \tag{4.36}$$

The probability that the particle scatters during the interval $(0,\tau)$ is

$$1 - \exp\left[-\tau W_T(E_{\underline{k}(\tau)})\right] \approx \tau W_T(E_{\underline{k}(\tau)}), \tag{4.37}$$

if $\tau \ll 1/W_T(E_{\underline{k}(\tau)})$.

We can reproduce the sequence of events shown in Fig. 4.3 by proceeding as follows[5]

[3] N.B.: if Γ is too large, too many self-scatterings may occur, slowing down the program. A reasonable choice is $\Gamma = 10^{15}$ s^{-1}. Scattering rates above this value do not satisfy the weak coupling approximation under which Fermi's Golden rule applies.

[4] A reliable source of random numbers is an essential building block for any Monte Carlo simulation code. As the output of any computer program is obviously entirely predictable, hence not truly random, the use of a computer to generate random numbers might seem a conceptual impossibility. Computer-generated "pseudo-random" numbers are actually obtained from one extremely long deterministic sequence of numbers, whose starting point (the "seed") can be chosen by the user. Different techniques are available to generate pseudo-random sequences. The Third Edition of Numerical Recipes warns us that the greatest lurking danger for a user today is that many out-of-date and inferior methods remain in general use (*sic*). Measuring the quality of this pseudo-random sequence is a nontrivial task. Among the accepted list of statistical checks devised to ferret out any nonrandomness in a sequence, are the battery of statistical tests named "diehard tests" by George Marsaglia. A more stringent assessment in the context of Monte Carlo device simulation is to verify that two different random number generators produce statistically the same results. If they don't, then at least one of them is not a good generator for your application program. In addition to statistical considerations, random number generators should be fast: the number of calls to these functions is of the order of millions in a typical Monte Carlo run – you get the idea.

[5] Such a "direct" simulation can actually be proven to be equivalent to the solution of the BTE in Chamber's path integral formulation.

1. Start with a particle in some initial $\underline{k}$-state and position $\underline{r}$ in real space. You may set $\underline{k} = 0$ and therefore $E_{\underline{k}} = 0$. If the simulation time τ_{sim} is long enough, the system will forget about the initial state and the final estimates (e.g., average velocities and energies) will not depend on this choice.

2. Select a flight time much smaller than the total scattering time, e.g., $\tau = 10^{-16}\,\text{s}$. Integrate the equations of motion for a free particle under the action of the electric field $\underline{\mathscr{E}}$ to obtain the change of the wavevector between collisions[1, Eq. (3.25)]

$$\Delta\underline{k} = -\frac{e\underline{\mathscr{E}}}{\hbar}\tau. \tag{4.38}$$

Notice that only the momentum component along the field is affected (x-direction in this laboratory), see the black segments in the left panel of Fig. 4.3. The change of the particle position along the field is

$$\Delta r = \langle v \rangle_\tau \tau \tag{4.39}$$

The "instananeous" carrier velocity $\langle v \rangle_\tau$ during the flight time τ is [1, Eq. (3.26)]

$$\langle v \rangle_\tau \approx \frac{1}{\hbar}\frac{\{\Delta E_{\underline{k}}\}_\tau}{\Delta k} = -\frac{\{\Delta E_{\underline{k}}\}_\tau}{e\mathscr{E}\tau}, \tag{4.40}$$

where Δk has been computed according to (4.38) and $\{\Delta E_{\underline{k}}\}_\tau$ is the increment of the carrier energy during τ. (In device Monte Carlo simulations, additional boundary conditions have to be included to account for particle absorption and injection at the contacts, specular or diffusive reflection at isolating boundaries, and scattering at heterointerfaces.)

3. Store all instantaneous quantities (this data will be used at the end of simulation to extract the desired transport properties by means of statistical estimators).

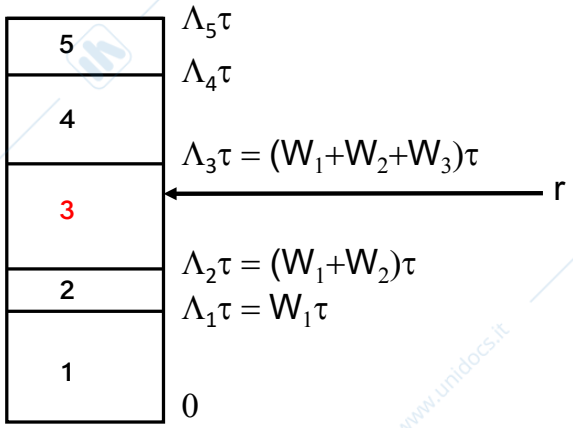
4. At the end of the flight, compute the scattering rates $W_j(E_{\underline{k}})$ (with $j = 1,\ldots,N$ running on the scattering mechanisms included in the simulation). Compute the total scattering rate $W_T = \sum_{j=1}^{N} W_j$ at the end of the free flight; pick a random number r uniformly distributed in the interval $[0,1]$. A scattering occurs if $r < \tau W_T$. If a scattering event is detected, then select one mechanism using the discrete inversion technique described in Appendix C, that is choose the nth-scattering mechanism if

$$\Lambda_{n-1}(E_{\underline{k}})\tau < r < \Lambda_n(E_{\underline{k}})\tau, \tag{4.41}$$

where r is a random number lying between zero and one, and the cumulative sum $\Lambda_n(E_{\underline{k}})$ is defined as

$$\Lambda_n(E_{\underline{k}}) = \sum_{j=1}^{n} W_j(E_{\underline{k}}). \tag{4.42}$$

If $r > \Lambda_N(E_{\underline{k}})\tau = W_T(E_{\underline{k}})\tau$, no scattering has occurred and the state of the particle is unchanged. Example ($N = 5$):



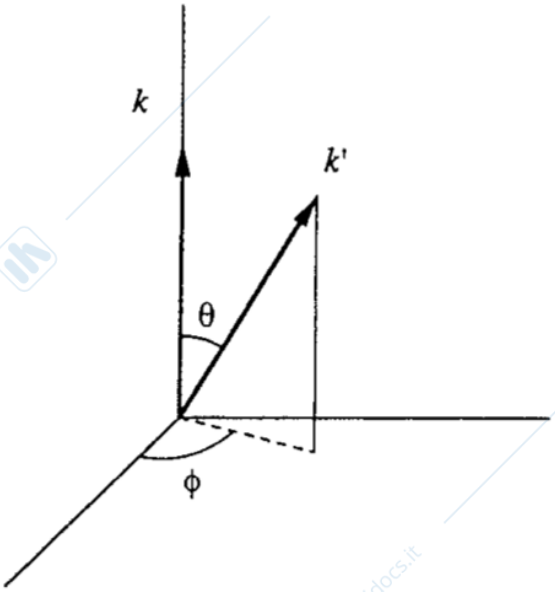


Figure 4.4.: Local reference system for the final states $\underline{k}'$ with polar axis parallel to the initial state $\underline{k}$. Reproduced from [1, Fig. 2.2].

5. Once the scattering mechanism that caused the end of the electron free flight has been determined, the new electron state $(\underline{k}', n')$ after scattering must be chosen ($n' \neq n$ only for intervalley processes, and the position of the particle does not change since the scattering is considered instantaneous). The new state is selected according to the corresponding scattering probability $S(\underline{k}n, \underline{k}'n')$ in the collision term (1.3).[6] When updating $\underline{k}$, it is convenient to work with the local reference frame in which the polar axis is parallel to the initial wavevector $\underline{k}$, see Fig. 4.4. The magnitude of $\underline{k}'$ is given by energy conservation (all energies are measured from the bottom of the valley; when an electron changes valley, as a result of an intervalley scattering process, the final energy should be computed as: $E_f = E_i \pm \hbar\omega_{if} - \Delta E_{fi}$, where Δ_{fi} is the energy separation between the valleys)

$$\gamma(E_\mathrm{f}) = \frac{\hbar^2 k'^2}{2m^*} \quad \rightarrow \quad k' = \sqrt{\frac{2m^*\gamma(E_\mathrm{f})}{\hbar^2}}, \tag{4.43}$$

and we need two random number to choose the scattering angles θ and ϕ as defined in Fig. 4.4. All the scattering mechanisms considered here are isotropic with respect to the azimultal angle ϕ, so that we can choose $\phi = 2\pi r_\phi$ where r_ϕ is a uniform random number lying between 0 and 1. The polar angle θ between $\underline{k}$ and $\underline{k}'$, can be selected stochastically by equating the (normalized) cumulative distribution function of the scattering angle with a uniform random number r_θ lying between 0 and 1

$$\frac{W(\theta)}{W(\underline{k})} = \frac{\int_0^\theta P(\theta')\mathrm{d}\theta'}{W(\underline{k})} = r_\theta, \tag{4.44}$$

where $P(\theta)\mathrm{d}\theta$ is the probability of scattering into the angles between θ and $\theta + \mathrm{d}\theta$. In some simple cases, the integral (4.44) can be inverted analytically.[7] Examples are: (i) impurity scattering [1,

[6] We assume that carrier densities are low enough so that Pauli's exclusion principle may be neglected, i.e., we omit the factor $(1 - f(\underline{k}'))$ that appears in the collision integral of the BTE. Phase space filling effects may be included within single-particle simulations through the self-scattering rejection technique proposed by Bosi and Jacoboni, later extended by Lugli and Ferry for ensemble simulations.

[7] When direct inversion methods are not applicable, the scattering angle θ can be determined with the rejection technique, which applies to any bounded probability distribution function (PDF) in a finite interval, and it is basically like throwing darts at a graph of the PDF, see Appendix C.

Eq. 2.54]

$$\cos\theta = 1 - \frac{2r_\theta}{1+(1-r_\theta)\left(\frac{2k}{q_0}\right)^2}, \tag{4.45}$$

and (ii) polar optical scattering [1, Eq. (2.106)]

$$\cos\theta = \frac{(1+f)-(1+2f)^{r_\theta}}{f} \tag{4.46}$$

with

$$f = \frac{2\sqrt{E_\mathrm{i}E_\mathrm{f}}}{(\sqrt{E_\mathrm{i}}-\sqrt{E_\mathrm{f}})^2}. \tag{4.47}$$

Intra- and inter-valley deformation potential scattering are isotropic, which means that θ can be selected as [1, Eq. 3.11]

$$\cos\theta = 1 - 2r_\theta. \tag{4.48}$$

Having computed the scattering angles in the local reference frame of Fig. 4.4, the components of $\underline{k}'$ can be rotated back to the original laboratory frame by means of the matrix transformation [1, Eq. (3.20)]

$$\begin{pmatrix} k'_x \\ k'_y \\ k'_z \end{pmatrix} = \begin{pmatrix} \frac{k_y}{\sqrt{k_x^2+k_y^2}} & \frac{k_x k_z}{k\sqrt{k_x^2+k_y^2}} & \frac{k_x}{k} \\ \frac{-k_x}{\sqrt{k_x^2+k_y^2}} & \frac{k_y k_z}{k\sqrt{k_x^2+k_y^2}} & \frac{k_y}{k} \\ 0 & \frac{-\sqrt{k_x^2+k_y^2}}{k} & \frac{k_z}{k} \end{pmatrix} \begin{pmatrix} k'\sin\theta\cos\phi \\ k'\sin\theta\sin\phi \\ k'\cos\theta \end{pmatrix}. \tag{4.49}$$

6. The steps above are repeated until the time variable t has exceeded the simulation time τ_sim.

The flowchart of a single particle Monte Carlo simulation is shown in Fig. 4.5.

At the end of the MC simulation, the relevant carrier transport properties can be extracted in the followig way. In stationary and homogeneous conditions, the time average of a quantity $A(\underline{k}(t))$ during the simulation time τ_sim as

$$\langle A\rangle_{\tau_\mathrm{sim}} = \frac{1}{\tau_\mathrm{sim}}\int_{\tau_\mathrm{sim}} \mathrm{d}t' A(\underline{k}(t')) = \frac{1}{\tau_\mathrm{sim}}\sum_i \int_{\tau_i} \mathrm{d}t' A(\underline{k}(t')) = \frac{1}{\tau_\mathrm{sim}}\sum_i \tau_i \langle A\rangle_{\tau_i} \tag{4.50}$$

where the integral over τ_sim has been split into the sum of integrals over all free flights of duration τ_i. For example, remembering the expression (4.40) for the instantaneous velocity along the field

$$\langle v\rangle_\tau = -\frac{\{\Delta E_{\underline{k}}\}_\tau}{e\mathscr{E}\tau}, \tag{4.51}$$

the mean velocity estimated at time t can be obtained by accumulating the variable $\{\Delta E_{\underline{k}}\}_\tau$ over each flight time up to time t

$$\langle v\rangle_t = -\frac{\sum_i \{\Delta E_{\underline{k}}\}_{\tau_i}}{e\mathscr{E}t}. \tag{4.52}$$

Likewise, the average carrier energy at the current time t is [1, Eq. (3.28)]

$$\langle E\rangle_t = \frac{1}{t}\sum_i \tau_i \langle E\rangle_{\tau_i}, \tag{4.53}$$

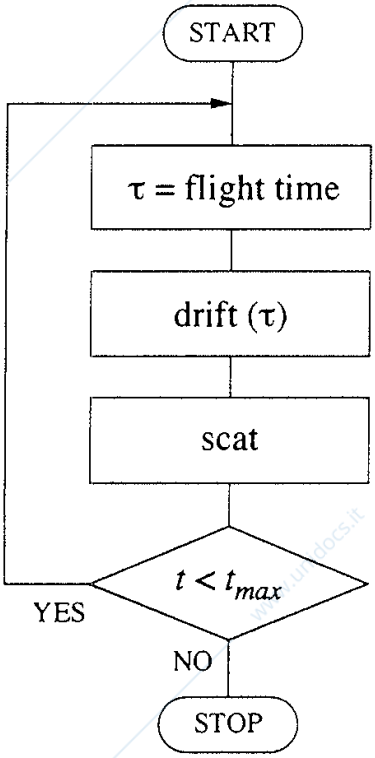


Figure 4.5.: Flowchart of a single particle Monte Carlo simulation. Reproduced from [1].

with $\langle E\rangle_\tau \approx (E_i + E_f)/2$, E_i and E_f being the energies at the begininning and at the end of the free-flight, respectively. See Fig. 4.6 for a convergence analysis.

In a similar way the electron distribution function can be obtained: during the MC run, the instantaneous wavevectors of the particle are recorded; at the end of the simulation, a mesh of energy points is set up; the time spent by the sample electron in each cell of the mesh is computed with the Matlab command *hist*. For large τ_{sim}, this time, conveniently normalized, represents the energy distribution function. Fig. 4.7 shows a comparison between the MC distribution and the Maxwellian fit (the density of states times the Maxwell-Boltzmann distribution)

$$f_M(E) \propto \sqrt{E} e^{-\frac{E}{k_B T_e}}, \tag{4.54}$$

where the fitting parameter T_e can be interpreted as the electron temperature. At low fields, the distribution function should approach the analytical fit with $T_e = 300\,\text{K}$. At higher fields, when intervalley scattering becomes possible, the distribution function becomes bimodal; in this case the analytic approximation (4.54) fails.

4.2.1. Velocity-field curve

Following the steps above, implement a single-particle non-selfconsistent Monte Carlo code to analyze carrier transport in bulk GaAs, assuming a two-parabolic-band model including the Γ and L valleys, with the following scattering mechanisms included: polar optical scattering, elastic acoustic scattering, intervalley deformation potential scattering, and impurity scattering.

1. Set the electric field to 0.1 kV/cm and verify that T_e is close to the lattice temperature.

2. Plot the velocity-field curve by repeating the Monte Carlo simulation for different electric fields. For each value of the applied electric field, verify the convergence of the Monte Carlo simulation

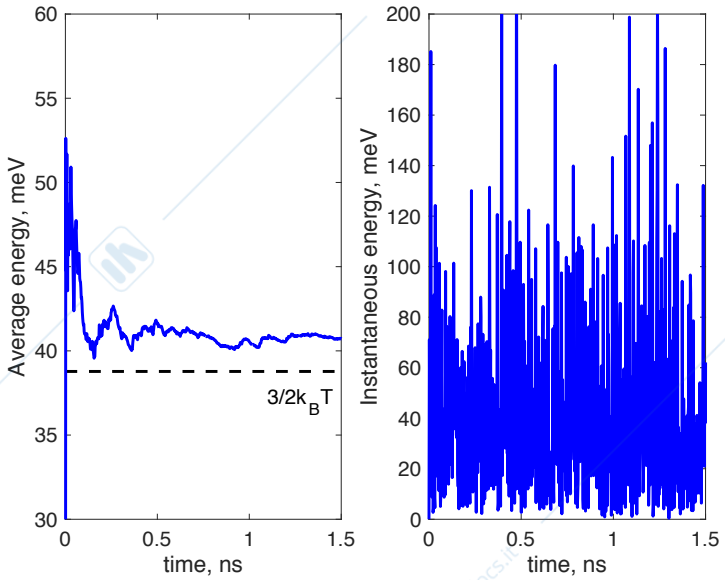


Figure 4.6.: (Left) The energy estimator $\langle E \rangle_t$ defined in (4.53) as a function of time (solid line) obtained from a single-particle Monte Carlo simulation at $\mathscr{E} = 1\,\mathrm{kV/cm}$ and $T = 300\,\mathrm{K}$. Also shown is the theoretical limit at zero field, $\frac{3}{2}k_B T$ (dashed line). The convergence of the estimator depends on simulation time, temperature, and electric field. (Right) The corresponding instantaneous energy of the particle in the same Monte Carlo run. It is important verify the convergence of all quantities, don't commit a "statistical sin"! Eric Grimson, John Guttag, and Ana Bell. 6.0002 Introduction to Computational Thinking and Data Science. Lecture 15: Statistical Sins and Wrap Up. Fall 2016. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

by plotting the estimated carrier velocity as a function of time and verify that it lies within a confidence interval upon convergence. You should observe a decrease of the electron velocity when the electric field is increased above the threshold value ($\approx 4\,\mathrm{kV/cm}$), an effect known as negative differential resistivity (NDR), due to the transfer of electrons from the Γ-valley to the L-valleys, where they have a larger effective mass,[8] see Fig. 4.8, upper panel.

3. For each electric field check the average energy, which should be converge close to $\frac{3}{2}k_B T$ for low fields, see Fig. 4.6. Compare the results with Fig. 4.8, lower panel.

4. Plot the average occupation of the Γ and L valleys as a function of the electric field, see [1, Fig. 3.16]. N.B.: there are four equivalent L valleys in the Brillouin zone; effective masses, intervalley coupling constants and intervalley phonon energies are reported in [1, Appendix A].

Here is a Matlab template of a single-particle Monte Carlo code (please use units complying with the International System for all quantities). The template is available in the material course folder (*MC_template.m*). To avoid confusion, let's use the suffixes i, f, and s to indicate energies before the

[8] The NDR phenomenon (in part also caused by band nonparabolicity) is at the basis of the Gunn effect. This effect consists in very fast current oscillations, (in the range of 10-30 GHz) obtained with the application of a constant voltage of high enough intensity to samples of materials, such as GaAs and InP, that present NDM. Obviously, such an effect is of great practical importance for the generation of microwaves. The reason for this phenomenon is well understood: NDR implies that a local fluctuation of charge density tends to grow with time, and this brings about the formation of the high-field domains that lower the current. The domains travel with the drift velocity of the electrons, which is the same inside and outside the domain, owing to the shape of the velocity-field curve. When they reach the anode they disappear, and the current increases until a new domain is formed. The period of the Gunn oscillations is therefore given by the transit time of the electrons across the sample and can be very short.

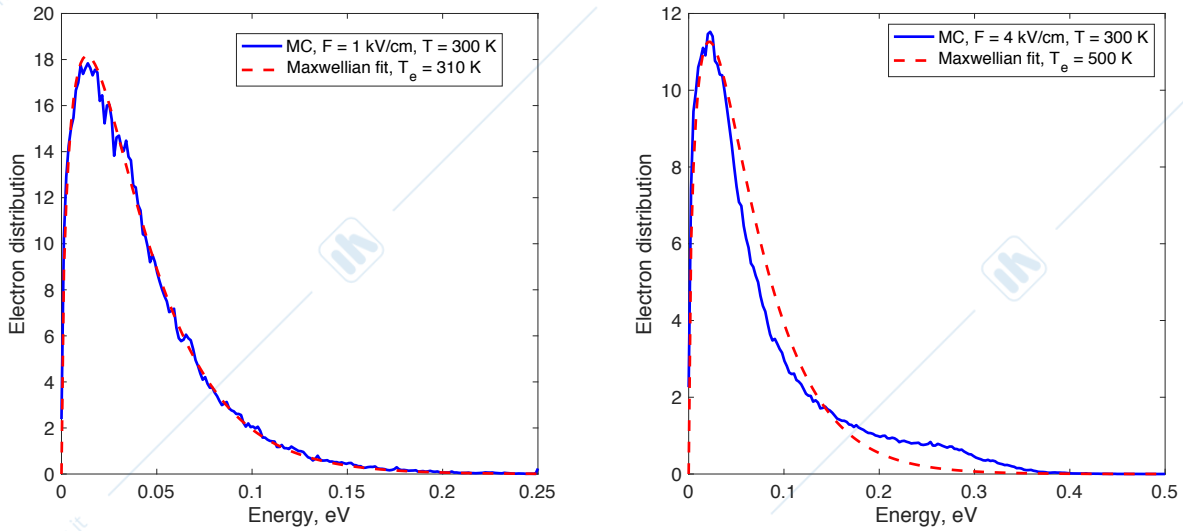


Figure 4.7.: Left. Energy distribution computed with the MC approach (solid line) at $F = 1\,\mathrm{kV/cm}$ and $T = 300\,\mathrm{K}$ in bulk GaAs, and its Maxwellian fit (dashed line) with $T_e = 310\,K$. Right. Energy distribution computed with the MC approach (solid line) at $F = 4\,\mathrm{kV/cm}$ and $T = 300\,\mathrm{K}$ in bulk GaAs, and its Maxwellian fit (dashed line) with $T_e = 500\,K$. The MC simulation time was $\tau_{\mathrm{sim}} = 5\,\mathrm{ns}$.

drift, at the end of the drift (just before the scattering), and after the scattering, respectively. For example, *Ekini* is the kinetic energy (measured from the bottom of the valley where the particle is) before the drift, and so on. The prefix *v* indicates vector quantities used to store particle data as a function of time.

N.B.: before activating intervalley scattering, verify that the code works for a single valley. For this reason the code corresponding to intervalley scattering has been commented below.

```
Q = 1.6021766208e-19; % elementary charge, C
T = 300; % temperature, K
tau_sim = 1e-9; % simulation time, s
tau = 1e-16; % flight time, s
Fx = 1e4; % electric field along the x-axis, V/m (0.1 kV/cm)
time = 0:tau:tau_sim; % time axis, s
nt = length(time); % # of time steps
E0 = [0 0.32]*Q; % energy reference of G and L valleys, J
%
% initial state
iv = 1; % valley index (1 -> G, 2 -> L)
kx = 0; ky = 0; kz = 0; % wavevector, 1/m
%
% statistical data
vEkini = zeros(1,nt); % array of kinetic energies before drift
vEkinf = zeros(1,nt); % array of kinetic energies after drift
...
%
%
% Monte Carlo main loop ------------------------------------------
for ii = 1:nt % time loop
```

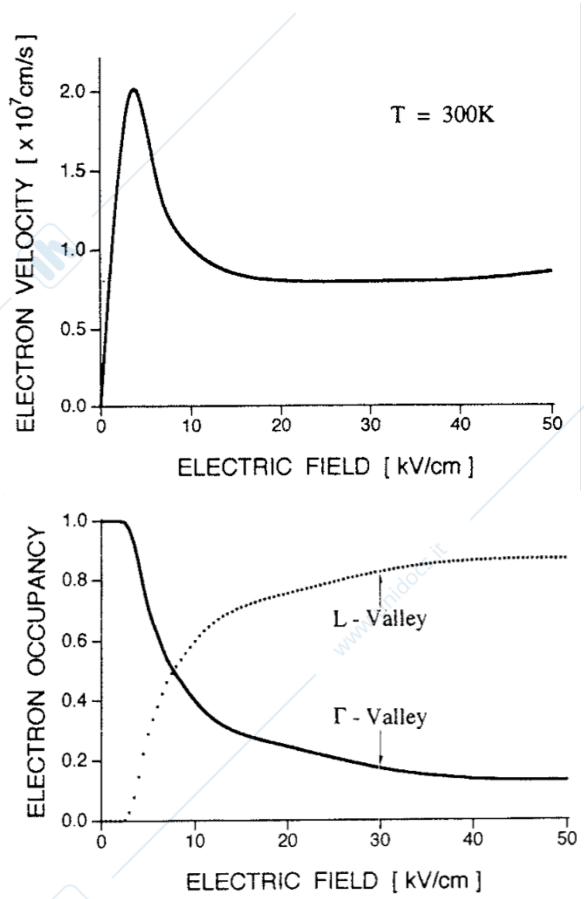


Figure 4.8.: Electron velocity and electron occupancy in the Γ- and *L*-valleys versus electric field in bulk GaAs. Reproduced from [1].

```
t = time(ii); % current time
%
% update energy
Ekini = k2engy(kx,ky,kz,iv); % kinetic energy before drift, J
Etoti = Ekini + E0(iv); % total energy (kinetic + potential), J
%
% drift particle in momentum space
kx = kx - Q*Fx*tau/HBAR;
Ekinf = k2engy(kx,ky,kz,iv); % kinetic energy after drift, J
Etotf = Ekinf + E0(iv); % total energy (kinetic + potential), J
% drift particle in real space
...
% collect statistics
vEkini(ii) = Ekini;
vEkinf(ii) = Ekinf;
...
% compute scattering rates
[Wpop_em,Wpop_ab] = pop_scat(Ekinf,T,iv);
[Waco]            = aco_scat(Ekinf,T,iv);
[Wimp]            = imp_scat(Ekinf,T,iv);
% [Wiv_em Wiv_ab] = iv_scat(Ekinf,T,iv);
W = [Wpop_em Wpop_ab Waco Wimp]; % scattering rates, 1/s
```

```
...
% choose scattering mechanim and final state
L = cumsum(W);
r = rand; % pick a random number
%
if(r < L(1)*tau) % select polar optical, emission
Ekins = Ekinf - hwpop; % kinetic energy after scattering, J
knorm = engy2k(Ekins,iv);
% compute scattering angles theta and phi
...
% rotate back to laboratory frame (anisotropic scattering only)
Rx = [ ky/sqrt(kx^2+ky^2), ...
       kx*kz/(sqrt(kx^2+ky^2+kz^2)*sqrt(kx^2+ky^2)),  ....
       kx/sqrt(kx^2+ky^2+kz^2)];
Ry = [-kx/sqrt(kx^2+ky^2), ...
       ky*kz/(sqrt(kx^2+ky^2+kz^2)*sqrt(kx^2+ky^2)), ...
       ky/sqrt(kx^2+ky^2+kz^2)];
Rz = [ 0, ...
      -sqrt(kx^2+ky^2)/sqrt(kx^2+ky^2+kz^2), ...
       kz/sqrt(kx^2+ky^2+kz^2)];
kve = knorm*[sin(theta)*cos(phi); sin(theta)*sin(phi); cos(theta)];
kx = Rx * kve; ky = Ry * kve; kz = Rz * kve;
%
elseif(r < L(2)*tau) % select polar optical, absorption
...
elseif(r < L(3)*tau) % select acoustic scattering
...
elseif(r < L(4)*tau) % select impurity scattering
...
% elseif(r < L(5)*tau) % select intervalley scattering, emission
% iv = 1.*(iv==2) + 2.*(iv==1); % update valley index
...
% elseif(r < L(6)*tau) % select intervalley scattering, absorption
...
else; end % no scattering has occurred (do nothing)
end % Monte Carlo main loop ---------------------------------------
%
%
% post-processing: drift velocity, diffusivity, et alia -------------
% Some quantities needed in the velocity and energy estimators:
vDeltaEkin =  vEkinf - vEkini; % energy variation during drift, J
vEkin_tau  = (vEkini + vEkinf)/2; % average energy during drift, J
...
```

Some useful functions:

```
% energy dispersion relation, (kx,ky,kz) -> E
k2engy = @(kx,ky,kz,iv) (iv==1)*HBAR^2*(kx^2+ky^2+kz^2)/(2*meff_G) + ...
                        (iv==2)*HBAR^2*(kx^2+ky^2+kz^2)/(2*meff_L);
%
% inverse dispersion relation E -> norm(kx,ky,kz)
engy2k = @(E,iv) (iv==1)*sqrt(2*meff_G*E/HBAR^2) + ...
                 (iv==2)*sqrt(2*meff_L*E/HBAR^2);
```

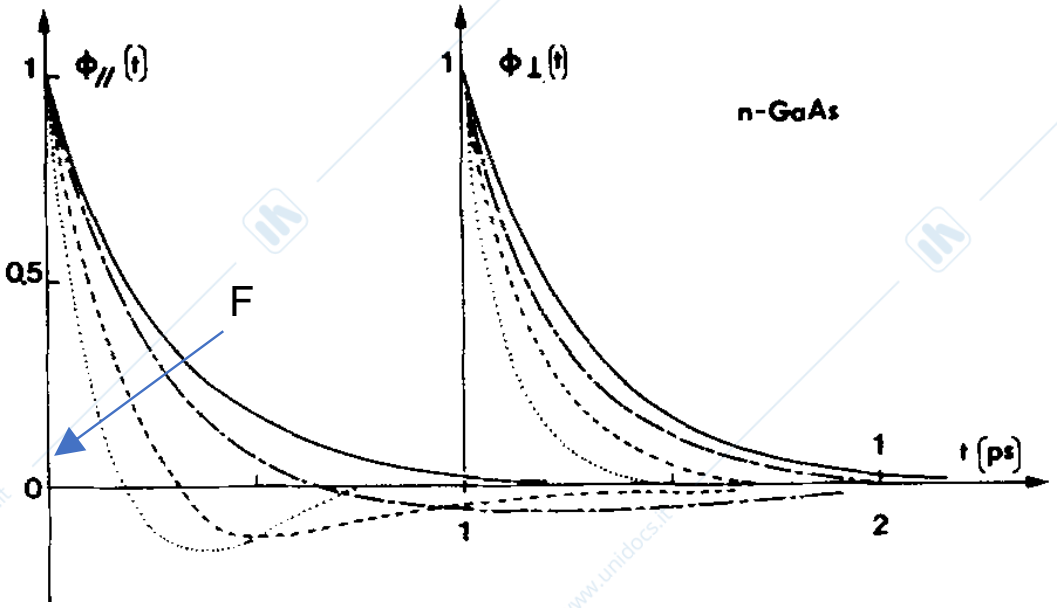


Figure 4.9.: (Normalized) autocorrelation function of the longitudinal $\parallel$ and perpendicular $\perp$ velocity fluctuations for increasing values of the electric field (F = 0 kV/cm, 5 kV/cm, 10 kV/cm, and 20 kV/cm) in GaAs. Reproduced from [36].

N.B. that in a typical Monte Carlo run, very unlikely events are bound to happen. Keep your code as simple as possible, premature optimization is never a good idea.[9] Consider all possibilities; check that the particle remains in the energy range in which the scattering rates are computed. If not, that probably means that the particle is not scattering properly or that the electric field is too high (please note that the validity of our model is restricted to relatively low electric fields where impact ionization processes are negligible and our description of the electronic structure in terms of "valleys" makes sense). Make sure that optical phonon emission is not allowed below the optical phonon energy (set the corresponding scattering rate to zero if the initial energy of the electron is below $\hbar\omega_{\mathrm{LO}}$). You might need to follow the history of the particle to make sure that the code is doing what it is supposed to do. Matlab offers powerful debugging tools. You may set breakpoints to pause the execution of the MATLAB code to examine the value or variables where you think a problem could be. To add a standard breakpoint in the Editor, click the breakpoint alley at an executable line where you want to set the breakpoint. The breakpoint alley is the narrow column on the left side of the Editor, to the right of the line number. The prompt in the Command Window changes to $K\gg$ indicating that MATLAB is in debug mode and that the keyboard is in control. MATLAB pauses at the first breakpoint in the program. In the Editor, a green arrow just to the right of the breakpoint indicates the pause. The program does not execute the line where the pause occurs until it resumes running. You can end the debugging session by clicking Quit Debugging in the Editor tab.

4.2.2. Diffusivity

Monte Carlo simulations can be used to derive important transport parameters needed by simplified approaches, as drift-diffusion and hydrodynamic models. For example, within the drift-diffusion approximation, both carrier mobility and diffusivity must be introduced as a function of the local electric

[9] The Yoda of Silicon Valley, Donald Knuth, master of algorithms, reflects on 50 years of his opus-in-progress, "The Art of Computer Programming". By Siobhan Roberts. New York Times, 2018.

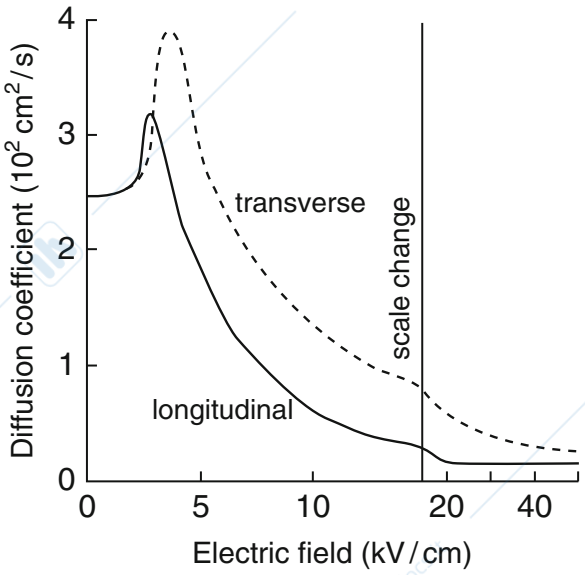


Figure 4.10.: Comparison between longitudinal and transverse diffusivities obtained in with MC simulations in bulk GaAs. Reproduced from [2, Chapter 15].

field. Einstein relation relates the diffusion coefficient D with the mobility μ of the carriers

$$D = \mu k_B T / e, \tag{4.55}$$

but it is strictly valid only at equilibrium. Although several (more or less empirical) models relating the carrier mobility to the electric field are available to accurately reproduce the DC and small-signal device behavior, much less information exists on the diffusion coefficient. The diffusion process is strictly related to noise.[10] The mathematical quantity that describes the common origin of diffusion and noise is the autocorrelation function of velocity fluctuations, see Fig. 4.9, which contains information on the magnitude of the fluctuations $\delta v(t) = v(t) - \langle v \rangle$ with respect to the average value $\langle v \rangle$ (drift velocity) and how they decay in time.[11] In principle, the diffusivity is a tensor quantity, but here we will restrict our analisys to diffusivity properties along the field direction, which allows us to use a simplified scalar notation where all the vector quantities are replaced with their *longitudinal* components. Following [2, Chapter 12], we define the *noise diffusivity* as

$$D = \int_0^\infty \mathrm{d}\tau C_\mathrm{v}(\tau) \tag{4.56}$$

where $C_\mathrm{v}(\tau) = \langle \delta v(t) \delta v(t+\tau) \rangle$ is the autocorrelation function of the velocity fluctuations. The physical meaning of (4.56) is that diffusion is related to the persistence of velocity fluctuations; in other words, if the autocorrelation function decreases rapidly to zero, its integral, i.e, the diffusivity is small. Suppose that the instantaneous velocities of the particle are recorded in the array *vel* and that the time step during the Monte Carlo run is $tau = 10^{-16}$ s. The autocorrelation function can be computed at the end of the MC simulation with the Matlab commands

[10] Two main types of fluctuations occur in semiconductor materials and, therefore, in devices: (i) fluctuations of carrier velocity or diffusion noise, associated to the Brownian motion of free carriers in the conduction or valence band, and (ii) fluctuations of carrier number or generation-recombination (GR) noise, due to transitions between conduction and valence bands, either direct or trap-assisted. These two phenomena are related, respectively, to intraband and interband scattering processes. Noise analisys in semiconductor devices can be carried out in the drift-diffusion approximation assuming small fluctuations and converting the linearized physical model into a Langevin equation, i.e., a partial differential equation complemented with a stochastic forcing term (the Langevin force) describing the microscopic noise sources [37]. On the other hand, noise is intrinsically included in any Monte Carlo model, since, by mimicking the physics of scattering, the method directly provides microscopic variables as realization of random processes, thus allowing to separate fluctuations from averages.

[11] Given a stochastic variable $A = \langle A \rangle + \delta A(t)$, the corresponding power spectral density is $S_A(\omega) = \int_{-\infty}^{+\infty} C_A(t) e^{\mathrm{i}\omega t} \mathrm{d}t$, where $C_A(t) = \langle \delta A(t) \delta A(0) \rangle$ is the corresponding autocorrelation function.

```
tau = 1e-16; % s
MAXLAG = 20000;
deltav = vel - mean(vel);
[Cv,lags] = xcorr(deltav,deltav,MAXLAG);
plot(lags*tau*1e12,Cv/length(deltav),'ro-')
set(gca,'FontSize',14,'FontName','Arial','box','on')
xlabel('time, ps')
ylabel('Autocorrelation function, m^2/s^2')
```

N.B.: a long simulation of few nanoseconds may be needed to obtain well converged results; in this case, *xcorr* may take to long, and you might consider increasing the time step in the calculation of the cross correlation, for example:

```
tau = 1e-15; % s
MAXLAG = 2000;
deltav = vel(1:10:end) - mean(vel);
[Cv,lags] = xcorr(deltav,deltav,MAXLAG);
plot(lags*tau*1e12,Cv/length(deltav),'ro-')
set(gca,'FontSize',14,'FontName','Arial','box','on')
xlabel('time, ps')
ylabel('Autocorrelation function, m^2/s^2')
```

At low fields, the autocorrelation function $C_v(\tau)$ shows a a typical exponential decay associated to velocity relaxation, i.e. it starts at some maximum positive value at $\tau = 0$ and then approaches zero asymptotically. At higher fields a negative tail associated to energy relaxation appears. That is, starting from some maximum positive value, C_v decreases rapidly, crosses zero, and reaches a minimum value less than zero, then returns toward zero asymptotically from the negative side. This is attributed to the competing effects of momentum and energy relaxation.

Plot the autocorrelation function of the longitudinal velocity fluctuations as a function of time. Compute the (longitudinal) noise diffusivity (4.56) as a function of the applied electric field, see Fig. 4.10. Compare the the low-field noise diffusivity with the diffusivity computed in the relaxation time approximation, see Section 4.1.

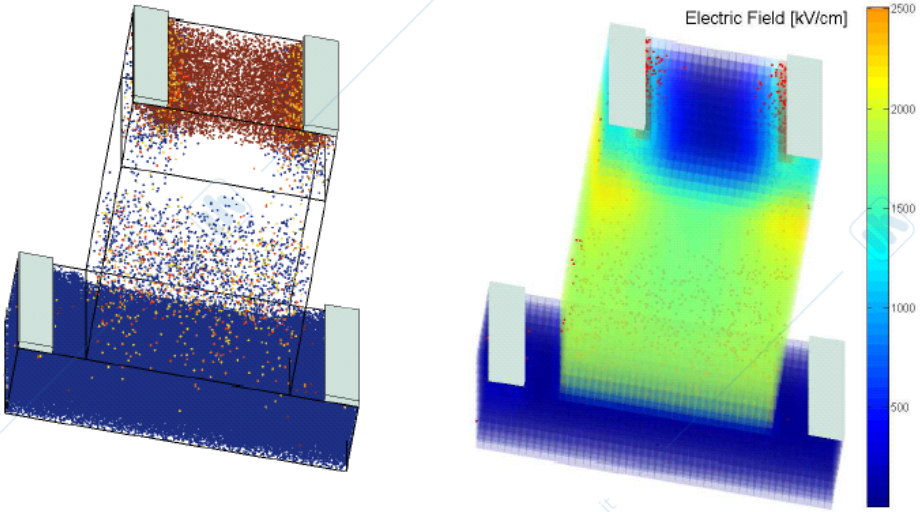
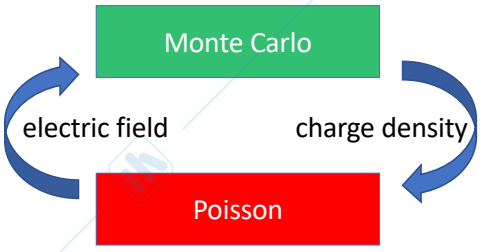


Figure 4.11.: Device Monte Carlo simulation of avalanche photodetectors.

4.3. Device Monte Carlo simulation

When the purpose of the analysis is the investigation of a steady-state, homogeneous phenomenon, as in section 4.2, it is sufficient in general to simulate the motion of a single electron; then, from ergodicity, we may assume that a sufficiently long path of this sample electron will give information on the behavior of the entire electron population. When, on the other hand, the transport process under investigation is not homogeneous and/or is not stationary, it is necessary to simulate a large number of electrons and follow them in their dynamic histories in order to obtain the desired information on the process of interest. Even though for steady-state analyses, one may simulate single particles from their entrances into the device to their exits, *ensemble* MC simulations, see Fig. 4.11, in which all considered carriers are simulated in parallel, are more often preferred. In this way, it is possible to analyze transients and take into account, when necessary, particle-particle interactions, including the exclusion principle.

The self-consistent simulation of semiconductor devices with MC techniques requires, as in the other methods, the periodic solution of Poisson's equation to obtain the spatial distributions of the electric field that moves the carriers.[12]



If a self-consistent solution of the Poisson's equation is desired, a few additional steps are necessary (particle mesh coupling)[13]

[12] It is thus necessary to solve Poisson's equation at very short time intervals during the simulation, to avoid that charge fluctuations act for an unphysical time, producing unphysical instabilities. For concentrations of the order of $10^{20}\,\mathrm{cm}^{-3}$ in silicon, as can be found in the drain region of a mosfet, a time-step for the application of the Poisson solver well below 10^{-15} s must be used, since the plasma period is of this order of magnitude.

[13] As any time-domain technique, the Monte Carlo approach is much less efficient, at least in its self-consistent version (when

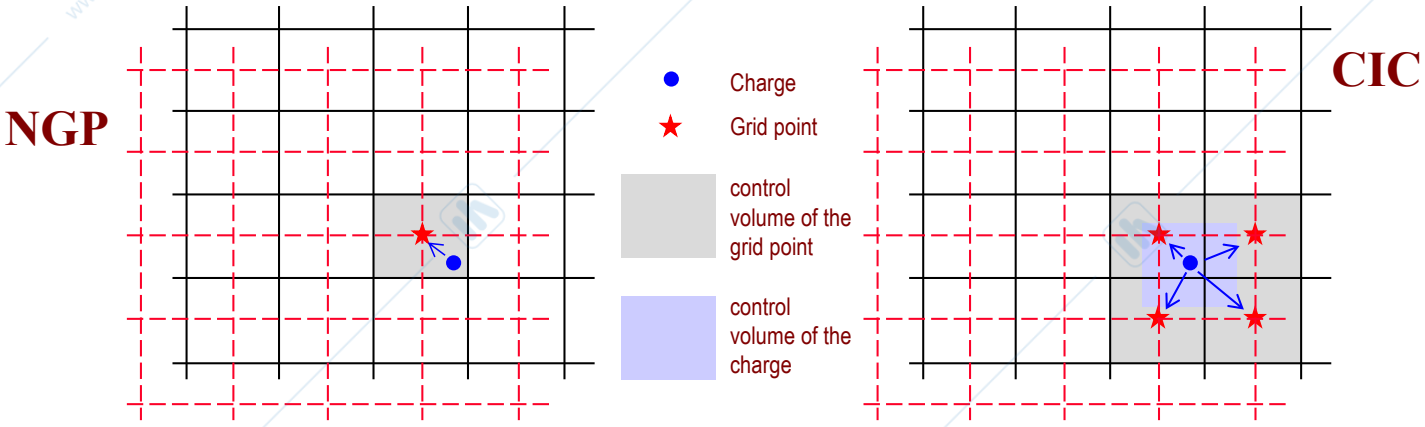


Figure 4.12.: Nearest-grid-point (NGP) and the cloud-in-cell (CIC) schemes.

1. Assign the charge of each particle to the mesh used to solve the Poisson equation (different charge assignments schemes are available)
2. Solve the Poisson equation in the mesh, by solving the associated finite element or finite difference discretized matrix problem.
3. Compute the components of the field on the nodes of the mesh.
4. Interpolate the field back to the particle positions.[14]

For this purpose, a grid must again be properly defined. The position of the charges, as described by the EMC algorithm is continuous, whereas Poisson's equation is solved on a mesh, hence, the charge associated with the individual particles must be mapped onto the mesh nodes in some fashion. The charge assignment and force interpolation schemes usually employed in self-consistent Monte Carlo device simulations are the nearest-grid-point (NGP) and the cloud-in-cell (CIC) schemes (this is done with the so-called *particle-mesh coupling*). Both technique assumes a discretization on finite elements

Poisson's equation is solved self-consistently with the charge distribution) if the time constants involved differ by orders of magnitude. This is the case e.g. in the simulation of bipolar devices, where intraband and interband (recombination) processes contribute with very different scattering times to determine the profile of the minority carriers. In addition to this limitation, the treatment of minority carriers (whose diffusion determines e.g. the current across a *pn*-junction) is problematic within a Monte Carlo framework, requiring statistical enhancement techniques that are quite cumbersome to implement; and in fact, most Monte Carlo studies of optoelectronic devices are restricted to the non-selfconsistent approximation or to unipolar devices. This does not imply that Monte Carlo device simulation is not viable in optoelectronics. On the contrary, it is probably the most general approach to the BTE, for its ability to incorporate physical models beyond what other techniques can handle (full-Brillouin-zone descriptions of the electronic structure, first principles phonon dispersions, wavevector-dependent scattering rates), limited only by the time one is willing to wait for a converged solution. As an example of how much physics can be included in a Monte Carlo code, we mention the pioneering works [38, 39, 40] on carrier-phonon and carrier-carrier scattering, with the aim to explain high-field carrier transport starting from the sole input of the ion pseudopotentials employed for the description of the electronic structure. Wavevector-dependent deformation potentials that rigorously satisfy selection rules along symmetry lines were derived from nonlocal empirical pseudopotential calculations of the electronic structure and density functional calculations of the lattice dynamics [40]. Their use shifted the empiricism of earlier Monte Carlo investigations (based on phenomenological deformation potential parameters) to a more fundamental level, leaving no fitting parameters involved, save for the initial choice for the pseudopotentials. Electron-phonon interactions and impact ionization were never treated in such detail within any other carrier transport formalism. Another example along these lines is the inclusion of interband tunneling within full-band Monte Carlo codes (by means of the time-dependent solution of the Schrödinger equation) to describe the coherent (between scattering events) drift of carriers at band crossing points [41].

[14] We have to make sure that the field due to a particle, when interpolated back to the position of the same particle, does not give rise to any unphysical force. The issue of the self-forces is particularly important in non-tensorial meshes.

implying grid points and control volumes. In the NGP scheme, the particle position is mapped entirely into the charge density at the closest grid point to a given particle. This has the advantage of simplicity, but leads to a noisy charge distribution, which may exacerbate numerical instability. Alternately, within the CIC scheme, a control volume is associated with each particle spanning several cells in the mesh, and a fractional portion of the charge per particle is assigned to neighboring grid points according to the fraction of the *cloud* overlapping with the relative cell volumes, see Fig. 4.12. This method has the advantage of smoothing the charge distribution due to the discrete charges of the particle-based method, but may result in an artificial *self-force* acting on the particle, particularly if an inhomogeneous mesh is used.[15] The standard NGP/CIC schemes can be summarized by the following four steps, (i) charge assignment to the mesh, (ii) solve Poisson's equation

$$\nabla \cdot (\varepsilon \nabla \phi) = \rho, \tag{4.57}$$

where ρ includes both mobile charges and fixed (doping) charges, (iii) calculate the force $E(r) = -\nabla\phi(r)$, and (iv) interpolate it to the particle positions. Poisson's equation must be augmented with appropriate Dirichlet boundary conditions (the potential value is known) at Ohmic and Schottky contacts, and Neumann (zero normal derivative of potential) on insulating or artificial boundaries.

Besides the boundary conditions of the Poisson's equation, another issue that has to be addressed in particle-based simulations is the real space boundary conditions for the particle part of the simulation.

1. Ohmic boundaries. Particles leaving the simulation region by crossing a device contact are marked as inactive; the event is recorded to estimate the contact current. Particles are injected periodically from an equilibrium distribution at Ohmic boundaries to maintain contact neutrality. The injection takes place in a uniform way along the cells adjacent to the contacts.

2. Schottky contacts: may be modeled by absorbing all electrons that hit the contact, but not injecting particles back into the device, or just the fraction of the distribution corresponding to energies exceeding the barrier height from the metal side.

3. Insulating surfaces. When particles hit insulating surfaces of the simulated region, reflecting or periodic conditions may be applied. To account for surface roughness, reflection at the boundaries of the device may be considered partially diffusive.

To simulate the steady-state behavior of a device, the system must be initialized in some initial condition with the desired potentials applied to the contacts, and then the simulation proceeds in a time-stepping manner until steady-state is reached. This process may take several picoseconds of simulation time and consequently several thousand time-steps based on the usual time increments required for stability. Clearly, the closer the initial state of the system is to the steady-state solution, the quicker the convergence. If one is, for example, simulating the first bias point for a transistor simulation, and has no a priori knowledge of the solution, a common starting point for the initial guess is to start out with charge neutrality, i.e., to assign particles randomly according to the doping profile in the device and based on the super-particle charge assignment of the particles, so that initially the system is charge neutral on the average. For two-dimensional device simulation, one should keep in mind that each particle actually represents a rod of charge into the third dimension. Subsequent simulations at the same device at different bias conditions can use the steady-state solution at the previous bias point as a good initial guess.

[15] Both NGP and CIC schemes may be used in semiconductor simulation subject to two restrictions, one, the permittivity must be constant, and, two, a tensor product mesh with uniform spacing in each axis direction must be employed. If these restrictions are ignored, self forces becomes nonzero and or spatial accuracy of forces is sacrificed. The two concepts are related. Self force refers to the force a charge may exert upon itself in an idealized situation namely a lone charge sitting in a domain of infinite extent and constant permittivity. Physically this lone charge should experience no force; numerically, any force experienced is self force. citeFischetti

4.3.1. Avalanche photodetectors

Never mind that the analysis of high-field carrier transport effects such as impact ionization, demands the use of full-Brillouin-zone realistic band structures, computed, e.g., by means of the empirical pseudopotential method of Lab. 1, impact ionization being a threshold process that occurs at energies typically far above the range of validity of the effective mass approximation. Never mind that impact ionization coefficients – the inverse mean distance an electron travels before impact ionizing – critically depend on both carrier-phonon and impact ionization scattering rates. That, besides the electronic structure, a realistic phonon dispersion relation is a critical piece of information needed for the description of carrier-phonon interaction beyond the simple deformation potential theory described in section 3.1.1. Never mind all those things. For the sake of simplicity, consider the two-valley MC model that you have used to compute the velocity field curve in GaAs, our conclusions will not be affected by these simplifications anyway. Consistently with the effective mass approximation, approximate the impact ionization rate with the popular Keldysh formula [38]

$$W_{\mathrm{ii}}(E) = \frac{P}{\tau_{\mathrm{ep}}(E_{\mathrm{th}})}\left(\frac{E-E_{\mathrm{th}}}{E_{\mathrm{th}}}\right)^2 \tag{4.58}$$

when $E < E_{\mathrm{th}}$, and zero otherwise, where E_{th} is the is some ionization-threshold energy, P is a coupling parameter related to the Coulomb matrix element, and $1/\tau_{\mathrm{ep}}$ is is either the carrier-phonon scattering rate at threshold, according to the original formulation by Keldysh, or the energy relaxation rate at threshold, according to Ridley's lucky drift theory. This expression has been used extensively in the past. The expression (4.58) determines the rate at which an impact ionization event will occur as a function of the initiating carrier's energy, once it has attained the threshold energy. The threshold energy is equal at least to band gap energy, and is typically much higher depending on the wavevector conservation condition. If the bands are all assumed to be parabolic, then a simple expression of the threshold energy can be found that depends on the band gap energy and the electron and hole effective masses. Here, however, we follow the usual approach of treating the parameters P and E_{th} as fitting quantities. The parameters proposed in [38] are $E_{\mathrm{th}} = 1.7\,\mathrm{eV}$ and $P/\tau_{\mathrm{ep}} = 2.5\times10^{15}\,\mathrm{s}^{-1}$.

Besides the impact ionization transition rate (4.58), we also need to know the rate at which carriers attain the threshold energy, i.e., the survival rate of high-energy carriers, because the initiating carrier must attain a relatively high energy before an impact ionization event can occur. This information can be obtained by means of a single particle MC simulation, in which the distances between two successive impact ionization events are recorded as the particle moves through the lattice under the effect of the electric field. The mean distance can be calculated after a sufficiently meaningful statistical sample has been gathered. The electron that initializes the impact ionization process is assigned an energy $E(\underline{k}) - E_{\mathrm{gap}}$, where E_{gap} is the energy of the semiconductor, and a random wavevector at this energy is selected. This choice of the final state after impact ionization scattering is consistent with the simplicity of (4.58). The inverse of the mean distance between two impact ionization events, also known as ionization coefficient, is shown in Fig. 4.13 as a function of the inverse electric field.

APDs exploit the avalanche multiplication of photogenerated carriers through impact ionization to amplify the detector current and improve the device sensitivity. A larger photocurrent (for the same illumination level) is obtained with respect to *pin* photodiodes, but, since impact ionization is fundamentally a random process, excess noise is generated along with the current gain due to random fluctuations in the timing of the impact ionization events as well as the number of impact ionization events, and thus the gain itself is randomized. This excess noise occurs in addition to the usual noise sources present in any semiconductor device. Although the device analysis shows that the responsivity increases with increasing avalanche multiplication, which, in turn, is larger if the hole and electron ionization coefficients are similar (i.e., if both carriers contribute to the avalanche process), large values of the multiplication factor are inconvenient because they increase noise and decrease speed. The device response is faster when only one type of carriers (e.g., electrons) trigger the ionization process. In fact, if both electrons and holes have the same ionization probability, an electron pulse injected into the high-field region generates holes which, instead of simply being collected in the p-side, in turn generate secondary electrons,

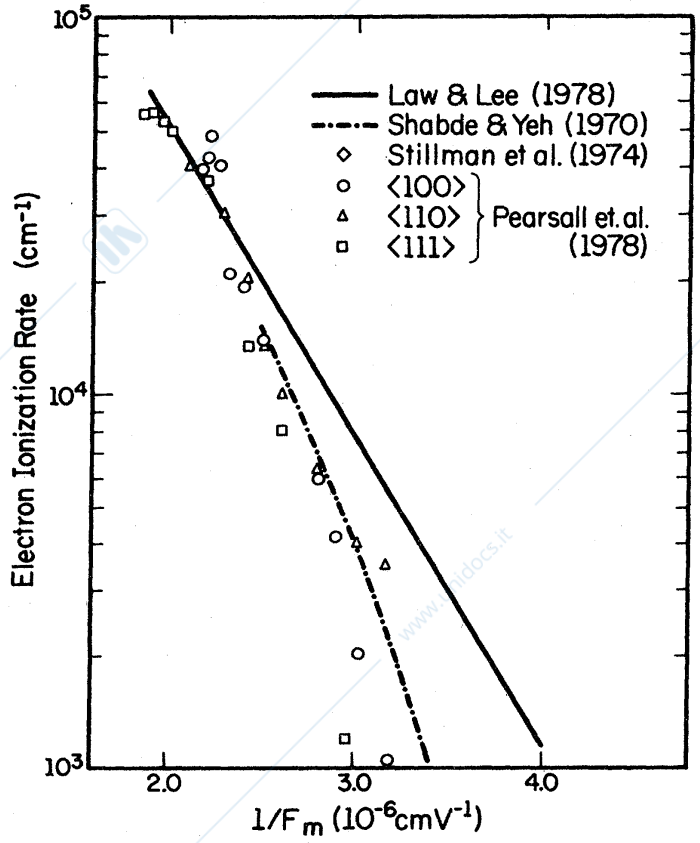


Figure 4.13.: Electron impact ionization coefficient as a function of the inverse electric field. Reproduced from [42].

whose delay before collection is of course larger than the delay of the electrons directly generated by the initial pulse. Secondary electrons generate new holes, and so on; if the multiplication factor is finite, the process finally dies out with a slow tail. Although a larger number of carriers is collected, such carriers are collected over a longer time interval, leading to an output current pulse with a large spread in time. This finally amounts to a slower response. Therefore, optimum noise and speed are achieved when the avalanche process is almost unipolar; as already mentioned, this happens when the ionization coefficient of the avalanche-triggering carrier is much larger than the ionization coefficient of the other carrier. For some materials, this condition is met naturally; it can be artificially introduced through bandstructure engineering and the use of superlattices.

The bandwith is proportional to the temporal response of the device. The carriers photogenerated outside of the depletion region must first diffuse to the depletion region edges, after which they are swept out of the device by the action of the reverse bias field (electrons drift towards the *n*-contact, holes towards the *p*-contact). The diffusion process is typically slow, and can take a relatively long time, depending on the width of the absorbing layer. Avalanche multiplication by impact ionization may occur in the high-field region. In order to absorb the incident radiation fully, the depletion region should be wide. However, as W increases, the transit time across the junction increases, thereby reducing the bandwidth of the device.[16]

Since APDs are inherently bipolar, their analysis in principle requires an ensemble Monte Carlo simulation of both electrons and holes. However, if we assume that holes do not initiate impact ionization

[16] In principle, the device includes two regions: the generation region (low to medium electric field) and the multiplication region (high field). The two regions can be physically the same, as in conventional APDs (see Fig. 4.14), or can be separated as in SAM-APDs (separate absorption and multiplication APDs) to afford the device designer the freedom to choose both the optimal absorption material and a suitable one-carrier-dominant avalanche region material [43].

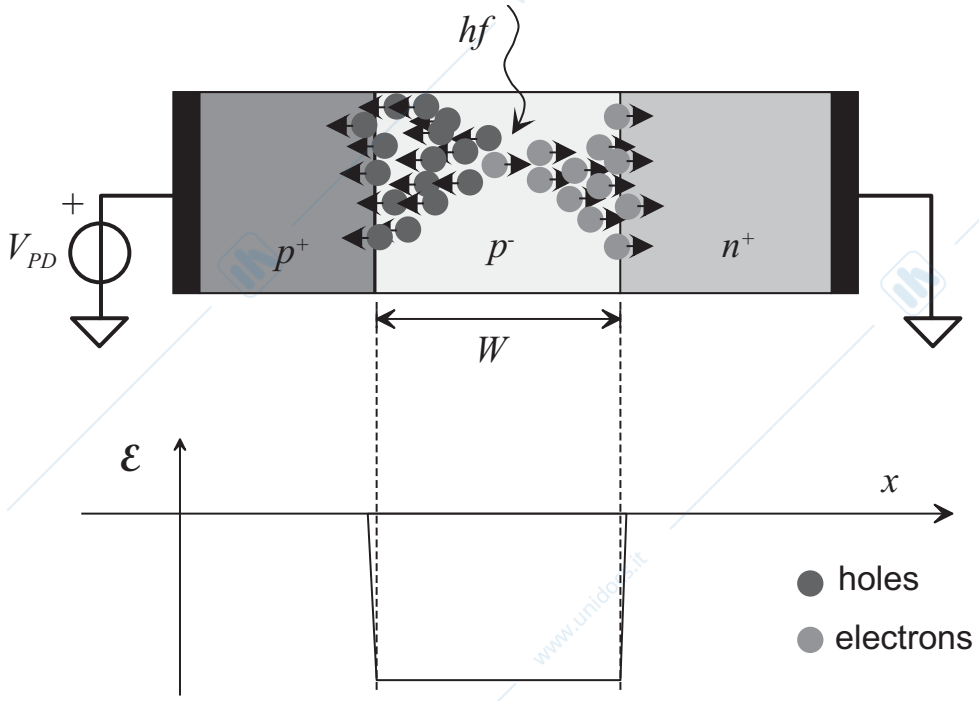


Figure 4.14.: Conventional *pin* APD structure with electric field profile. Reproduced from [43].

processes, and that the photogenerated carrier distribution does not contribute to the calculation of the electric field (low photocurrent limit), a non-selfconsistent unipolar analysis is appropriate.[17]

Let $Z(x)$ be the total number of electrons and holes generated by an electron injected at position x, including the initiating electron itself. Similarly define $Y(x)$ as the total number of electrons and holes generated by a hole injected at position x, including the hole itself. The total number of electron-hole pairs generated within the device as a result of the original carrier pair at location x, including the original electron-hole pair that initiated the entire multiplication process, is a random number $M(x)$ defined as

$$M(x) = \frac{Z(x) + Y(x)}{2}. \tag{4.59}$$

Our objective is to determine the statistics of M. If the electron-hole pair is injected at $x = 0$, the gain of the device is

$$G = M(0) = \frac{Z(0) + Y(0)}{2}. \tag{4.60}$$

Since a hole in $x = 0$ does not travel into the device and cannot ionize, $Y(0) = 1$, and $G = M(0) = (Z(0) + 1)/2$. The mean gain is therefore given by

$$\langle G \rangle = \frac{z(0) + 1}{2}. \tag{4.61}$$

[17] Several macroscopic approaches have been developed to compute the gain and noise characteristics of APDs [43, Sec. 4.11]. A fundamental assumption implicit in local models is that the probability of impact ionization by a carrier is independent of its ionization history, i.e., carrier multiplication is treated as a local process in which the ionization coefficients are functions only of the local value of the electric field. From a physical point of view, however, a newly generated carrier must travel some distance, the dead space, in order to build up sufficient energy to initiate an ionization. Therefore, the local ionization model overestimates the gain and noise figures in thin structures in which the dead space is a significant fraction of the multiplication region. History dependent models have been proposed to deal with the non-local nature of the impact ionization process [44]. Key aspects of non-local models are the concepts of dead space and the carrier enabled impact ionization coefficients. These enabled coefficients are defined as the ionization probability per unit distance experienced by carriers after travelling the dead space distance. Enabled ionization coefficients and dead space distances can be computed from Monte Carlo simulations. A rigorous account of dead space effects in the multiplication gain and noise properties of APDs can be obtained by using the recursive technique proposed by Hayat [45].

The excess noise factor F is defined as

$$F = \frac{\langle G^2 \rangle}{\langle G \rangle^2}. \tag{4.62}$$

The gain and the excess noise factor can determined from data collected from large number of trials. In other words, run several MC simulations starting with one electron at $x = 0$. If the electron impact ionizes, select the final state as described above, and simply add the secondary electrons to the MC simulation (place the generated secondary particles at the bottom of the conduction band). N.B.: rather than implementing an ensemble MC code, it is probably more convenient to simulate secondary particles by calling recursively a single particle MC program. The calculation continues until all the electrons are collected by the contacts. Particles should be removed from the simulation when their spatial coordinate x is found to be outside the simulation domain. If the number of particles keeps increasing, it probably means that the structure is in breakdown.

4.4. Conclusion

For didactical reasons, the Monte Carlo model proposed here is based on a minimal set of scattering mechanisms (ionized impurity scattering, polar optical scattering, intravalley and intervalley deformation potential scattering, impact ionization scattering) derived assuming a two-valleys model, with the objective to keep things as simple as possible, while still allowing for a realistic description of the velocity-field curve in the low-field regime and of the impact ionization coefficients in high-field conditions. Inclusion of additional scattering mechanisms (piezoelectric scattering, roughness scattering, alloy scattering if applicable, etc.) or lifting some of the approximations considered here would make the present analysis more accurate. For example, acoustic scattering could be modeled as inelastic (beyond the equipartition approximation, see section 3.1) to avoid artifacts in simulations at low fields, a more realistic description of the electronic structure along the lines of Lab. 1 could be employed. There is no limit to the physics that can be included in a Monte Carlo model. Poisson's equation could be solved self-consistently for complex geometries. The scattering kernels $S(\underline{k},\underline{k})$ could be rendered explicit functions of position and of time, functionals of the distribution function, and still the resulting BTE would be solved exactly (although in a statistical sense), without invoking approximations on the shape of the distribution function, of the electronic structure, and of the phonon dispersion relation; and without even requiring the models to be described analytically. What makes all this possible is the conceptually simple idea of using randomness for determinative purposes, which basically means following the random motion of carriers as they interact with the lattice and collecting statistics. And that's as close to the physics of carrier transport as you can get.

Appendix A.

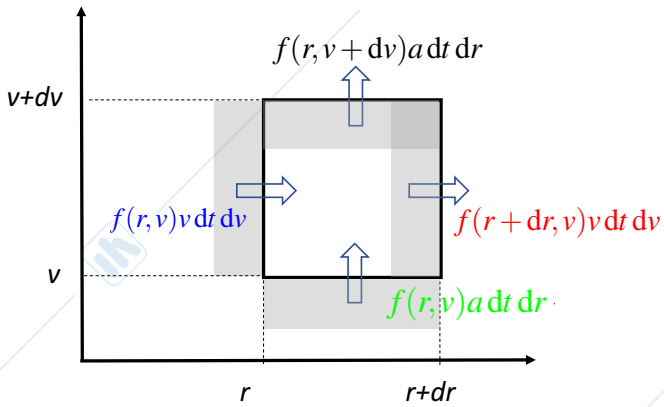
Elementary derivation of the BTE

Let us consider an infinitesimal region in the two-dimensional position-momentum space around the point (r,v). The number of particles in this elementary region

$$N = f(r,v,t)\,\mathrm{d}r\,\mathrm{d}v. \tag{A.1}$$

The variation with time of N is given by the balance between the numbers of particles that enter and exit from the faces of the region

$$\mathrm{d}N = f(r,v)v\,\mathrm{d}t\,\mathrm{d}v - f(r+\mathrm{d}r,v)v\,\mathrm{d}t\,\mathrm{d}v + f(r,v)a\,\mathrm{d}t\,\mathrm{d}r - f(r,v+\mathrm{d}v)a\,\mathrm{d}t\,\mathrm{d}r \tag{A.2}$$



or

$$\partial_t f = -v\partial_r f - a\partial_v f, \tag{A.3}$$

where a is the acceleration of the particle. Generalizing to a six-dimensional position-momentum space, one obtains the Vlasov equation of plasma physics (Liouville equation):

$$\dot{f} + \underbrace{\underline{v}\cdot\nabla_r f}_{\text{diffusion}} + \underbrace{\underline{a}\cdot\nabla_v f}_{\text{drift}} = 0, \tag{A.4}$$

which states that the total derivative of the phase-space distribution function is always equal to zero; the latter, in turn, contains a so-called diffusion term involving the real-space gradient of f, as well as a drift contribution involving the gradient of f in momentum space. In the case of an electron gas in a crystal, it is convenient to use the variables $\underline{r}$ and $\underline{k}$

$$\dot{f} + \underline{v}\cdot\nabla_r f + \underline{k}\cdot\nabla_{\underline{k}} f = 0. \tag{A.5}$$

Adding the effect of the collisions to the unperturbed dynamics gives the BTE (1.1)

$$\frac{\partial}{\partial t} f(\underline{r},\underline{k},t) + \underline{v}\cdot\nabla_r f + \underline{k}\cdot\nabla_{\underline{k}} f = \left.\frac{\partial f}{\partial t}\right|_{\text{coll}}. \tag{A.6}$$

Under suitable physical conditions, such purely classical transport theory may be successfully employed for the description of non-classical systems; this constitutes the essence of the semiclassical picture.

Appendix B.

Time-dependent perturbation theory

One of the main tasks of quantum mechanics is the calculation of transition probabilities from an initial state to another state under the influence of a time-dependent perturbation. Assume that the Hamiltonian can expressed as the sum of two terms

$$\hat{H}(t) = \hat{H}_0 + \lambda \hat{H}'(t), \tag{B.1}$$

where $\hat{H}_0$ is the "simple" (time-independent) part of the Hamiltonian, the part whose eigenvectors and eigenvalues can be solved exactly, and $\lambda \hat{H}'$ is the "complex" part that makes the problem intractable, λ being a dimensionless quantity that we have factored out of $\hat{H}'$. The basic idea of perturbation theory is to develop the unknown quantities (eigenvalues and eigenvectors of $\hat{H}$) in powers of λ, and to solve the resulting equations separately for the different orders. We will assume that the perturbation is weak enough to ensure convergence. The time-dependent Schrödinger equation we need to solve is

$$[\hat{H}_0 + \lambda \hat{H}'(t)]|\phi_n(t)\rangle = \mathrm{i}\hbar \frac{\partial}{\partial t}|\phi_n(t)\rangle. \tag{B.2}$$

Let's suppose that we are able to solve the unperturbed problem

$$\hat{H}_0|\psi_n\rangle = E_n|\psi_n\rangle. \tag{B.3}$$

It is convinient to use as basis the eigenstates of the unperturbed Hamiltonian including their time dependence

$$|\psi_n(t)\rangle = |\psi_n\rangle e^{-\mathrm{i}E_n(t-t_0)/\hbar}. \tag{B.4}$$

Inserting

$$|\phi_n(t)\rangle = \sum_n a_n(t)|\psi_n(t)\rangle \tag{B.5}$$

in (B.2), we have

$$[\hat{H}_0 + \lambda \hat{H}'(t)]\sum_n a_n(t)|\psi_n(t)\rangle = \sum_n [\mathrm{i}\hbar \dot{a}_n(t) + E_n a_n(t)]|\psi_n(t)\rangle, \tag{B.6}$$

or

$$\sum_n a_n(t)[\hat{H}_0 - E_n]|\psi_n(t)\rangle + \sum_n [a_n(t)\lambda \hat{H}'(t) - \mathrm{i}\hbar \dot{a}_n(t)]|\psi_n(t)\rangle = 0. \tag{B.7}$$

The first term vanishes. Multiplying on the left the second term by $\langle \psi_m(t)|$, integrating, and recalling the orthonormalization condition $\langle \psi_m|\psi_n\rangle = \delta_{mn}$, we have

$$\sum_n [a_n(t)\lambda H'_{mn}(t)e^{\mathrm{i}(E_m - E_n)(t-t_0)/\hbar} - \mathrm{i}\hbar \dot{a}_n(t)\delta_{mn}] = 0, \tag{B.8}$$

with

$$H'_{mn}(t) = \langle \psi_m|\hat{H}'(t)|\psi_n\rangle = \int \mathrm{d}\underline{r}\, \psi^*_m(\underline{r})H'(t)\psi_n(\underline{r}). \tag{B.9}$$

Defining

$$\hbar\omega_{mn} = E_m - E_n, \tag{B.10}$$

(B.8) may be rewritten as

$$\mathrm{i}\hbar\dot{a}_m(t) = \sum_n a_n(t)\lambda H'_{mn}e^{\mathrm{i}\omega_{mn}(t-t_0)}. \tag{B.11}$$

This equation is exact, and shows how the Hamiltonian generates the evolution of the system according to its matrix elements. At this point, we consider the perturbation expansion of the coefficients

$$a_m(t) = a_m^{(0)}(t) + \lambda a_m^{(1)}(t) + \lambda^2 a_m^{(2)}(t) + \ldots \tag{B.12}$$

and insert it in (B.11). Equating terms of like powers of λ on both sides of the resulting equation, we have for the first three terms

$$\dot{a}_m^{(0)}(t) = 0 \tag{B.13a}$$

$$\mathrm{i}\hbar\dot{a}_m^{(1)}(t) = \sum_n H'_{mn}(t)a_n^{(0)}(t)e^{\mathrm{i}\omega_{mn}(t-t_0)} \tag{B.13b}$$

$$\mathrm{i}\hbar\dot{a}_m^{(2)}(t) = \sum_n H'_{mn}(t)a_n^{(1)}(t)e^{\mathrm{i}\omega_{mn}(t-t_0)} \tag{B.13c}$$

The first of these equations shows that to zero order the state is constant in time, i.e., it remains in its initial condition: $a_m^{(0)}(t) = a_m^{(0)}(t_0) = a_m^{(0)}$. In particular, if we assume $a_n^{(0)} = \delta_{ni}$, i.e., that the system is in the unperturbed eigenstate $|\psi_i\rangle$ at time $t = t_0$, the first-order approximation $a_f^{(1)}(t)$ of $a_f(t)$ $(f \neq i)$ can be evaluated from (B.13b)

$$\mathrm{i}\hbar\dot{a}_f^{(1)}(t) = H'_{fi}(t)e^{\mathrm{i}\omega_{fi}(t-t_0)}. \tag{B.14}$$

To proceed further, we need to assume a given form for $H'(t)$. Let's assume that the perturbation H' depends harmonically on time

$$H'(t) = Fe^{-\mathrm{i}\omega t} + F^*e^{\mathrm{i}\omega t}. \tag{B.15}$$

This may readily be the case (the presence of both terms ensures the Hermiticity of the Hamiltonian), or these two terms could just be just the Fourier components of a more general (real) function of time. Integration of (B.14) from t_0 to t gives

$$a_f^{(1)}(t) = \frac{1}{\mathrm{i}\hbar}\int_{t_0}^{t}\left[F_{fi}e^{-\mathrm{i}\omega t'} + F_{fi}^*e^{\mathrm{i}\omega t'}\right]e^{\mathrm{i}\omega_{fi}t'}\mathrm{d}t' =$$

$$\frac{1}{\mathrm{i}\hbar}\left[\underbrace{F_{fi}\frac{e^{\mathrm{i}(\omega_{fi}-\omega)t} - e^{\mathrm{i}(\omega_{fi}-\omega)t_0}}{\mathrm{i}(\omega_{fi}-\omega)}}_{\text{absorption}(\omega_{\mathrm{f}}=\omega_{\mathrm{i}}+\omega)} + \underbrace{F_{fi}^*\frac{e^{\mathrm{i}(\omega_{fi}+\omega)t} - e^{\mathrm{i}(\omega_{fi}+\omega)t_0}}{\mathrm{i}(\omega_{fi}+\omega)}}_{\text{emission}(\omega_{\mathrm{f}}=\omega_{\mathrm{i}}-\omega)}\right] \tag{B.16}$$

It is clear that the first term corresponds to the absorption of a quantum of the perturbing field, and the second to the emission. Let us then assume that the final state has energy close to the one corresponding to an absorption or an emission, and keep only one term. We shall see that this approximation is in general reasonable, since the two terms become sharply peaked, and where one is essentially different from zero the other is practically zero. The probability of finding the system at time t in the final state $|\psi_f\rangle$ after an absorption is then

$$|a_f(t)|^2 = \frac{1}{\hbar^2}|F_{fi}|^2\frac{\left|e^{\mathrm{i}(\omega_{fi}-\omega)t} - e^{\mathrm{i}(\omega_{fi}-\omega)t_0}\right|^2}{(\omega_{fi}-\omega)^2} = \frac{1}{\hbar^2}|F_{fi}|^2\frac{\sin^2(\xi\Delta t)}{\xi^2\Delta t}\Delta t \tag{B.17}$$

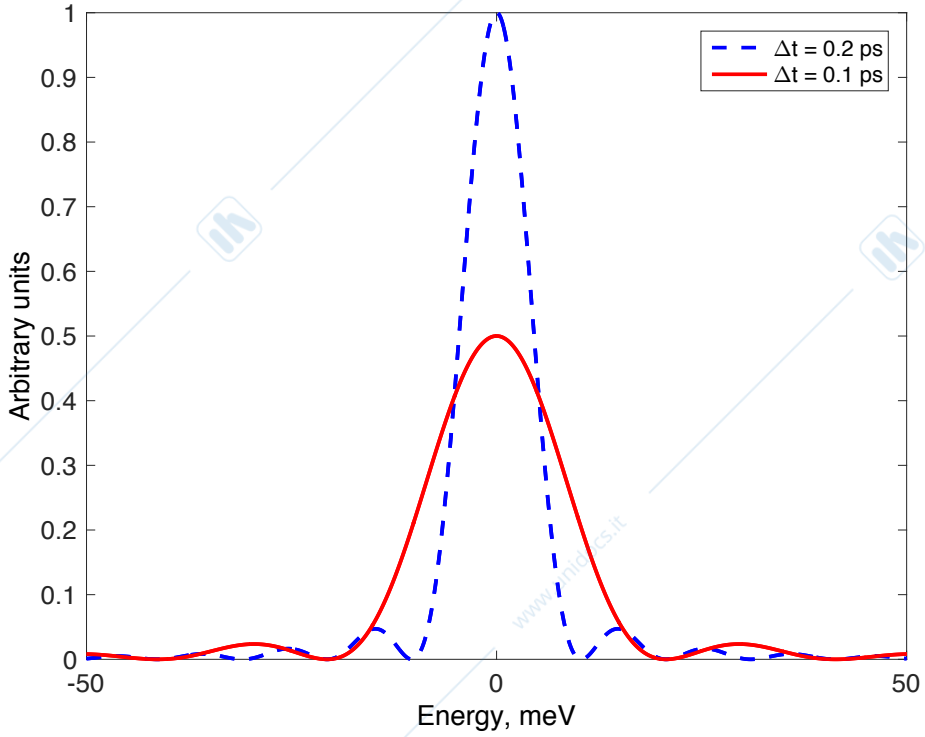


Figure B.1.: Shape of the function $\sin^2(\xi\Delta t)/(\xi^2\Delta t)$, with $\xi = E/(2\hbar)$, for different values of Δt. At increasing times, (B.18) approaches $\delta(\xi)$.

with $\xi = (\omega_{fi} - \omega)/2$ and $\Delta t = t - t_0$. The above formula gives the probability, to first order, of finding the system in the state $|\psi_f\rangle$, Δt seconds after it was prepared in the state $|\psi_i\rangle$, as a consequence of the absorption of a quantum of the perturbing field. Note that this probability is different from zero not only at exact energy conservation, but also close to it. A similar result is obtained for emission. The second-order probability can be obtained with similar calculations. It corresponds to two absorptions, or two emissions, or one absorption and one emission in the two possible time orders. For sufficiently long times Δt, we may approximate

$$\frac{\sin^2(\xi\Delta t)}{\pi\xi^2\Delta t} \approx \delta(\xi), \tag{B.18}$$

which gives

$$|a_f(t)|^2 = \frac{2\pi}{\hbar}|F_{fi}|^2\delta(E_f - E_i \mp \hbar\omega)\Delta t, \tag{B.19}$$

where the upper sign stands for absorption, and the lower sign for emission. The δ-function expresses the conservation of energy. (The fact that this function is strongly peaked justifies considering the two terms in (B.16) separately.) The probability in (B.19) is proportional to Δt. The transition probability per unit time can be defined as

$$S(i \to f) = \frac{2\pi}{\hbar}|F_{fi}|^2\delta(E_f - E_i \mp \hbar\omega), \tag{B.20}$$

which is the basic result of scattering theory known as Fermi's Golden rule: in order to compute the scattering rates appearing in the collision term of the BTE, you just need to define the perturbation Hamiltonian for each scattering mechanism, and compute its matrix elements between the unperturbed states of the system, which in our case are the Bloch's states of the crystal. Eq. (B.20) holds if we are in a weak (infrequent) scattering regime, so that we can take the limit $\Delta t \to \infty$ leading to energy conservation, without another scattering occurring.

Appendix C.

Generation of random numbers

The Monte Carlo procedure is based on the generation of random numbers. Any modern computer provides a routine for the generation of (pseudo)-random numbers evenly distributed over the interval $[0,1]$, i.e., starting from a given initial entry (called *seed*) the routine produces a periodic sequence of uncorrelated numbers uniformly distributed over $[0,1]$. Our objective is to generate random numbers according to a given distribution function.

C.1. Inversion method

Given a random variable U evenly distributed in $[0,1]$, and an invertible cumulative distribution function (cdf) $F(x)$, the random variable $X = F^{-1}(U)$ has cdf $F(x)$. This is the idea behind the inverse transform method, or simply the direct method. The algorithm proceeds as follows:

1. Generate a random number u uniformly distributed in $[0,1]$
2. Compute the value x such that $F(x) = u$
3. Take x to be the random number drawn from the distribution described by $F(x)$.

In order to prove this, recall that $F(x)$ is a non-negative and non-decreasing (monotone) function.

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x). \tag{C.1}$$

The last equality holds because $P(U \leq u) = u$ if U is uniform in $[0,1]$. For example, let's compute a random number uniformly distributed in the interval (a,b). According to the inversion technique we must find

$$u = F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' = \int_{a}^{x} \frac{\mathrm{d}x'}{b-a} = \frac{x-a}{b-a}, \tag{C.2}$$

which gives the intuitive result $x = u(b-a)+a$.

The inversion method has also a discrete version when $f(x)$ is the sum of δ functions. Consider a non-negative discrete random variable X with probability distribution function

$$P(X = k) = p_k, \tag{C.3}$$

with $k \geq 0$. Then the inversion technique gives

$$\begin{aligned} X &= 0 \qquad \text{if } U \leq p_0 \\ X &= k \qquad \text{if } \sum_{i=0}^{k-1} p_i < U < \sum_{i=0}^{k} p_i. \end{aligned} \tag{C.4}$$

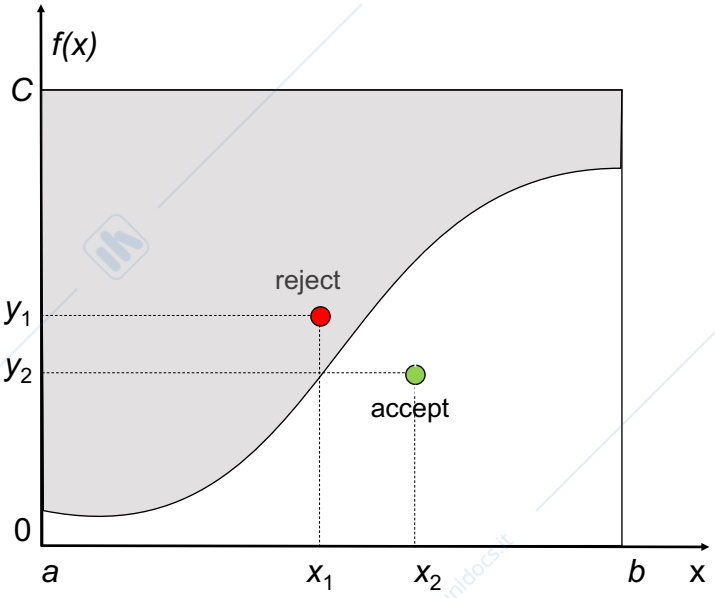


Figure C.1.: Rejection technique for a generation of a random number with probability density function $f(x)$ using random numbers evenly distributed in $[0,1]$.

C.2. Rejection method

The application of the inversion method is limited to one-dimensional problems for which the cumulative distribution and its inverse are known analytically. Consider a probability density function (pdf) $f(x)$ defined in interval $[a,b]$ and zero elsewhere; let C be a positive number such that $C \geq f(x)$ (it is not necessary to know the maximum of $f(x)$: the method works also with any overestimation of it). The algorithm, which is pretty much like throwing "darts" at the distribution, proceeds as follows:

1. Pick a random number r_1 uniformly distributed in $[0,1]$, then compute $x = a + (b-a)r_1$
2. Pick a random number r_2 uniformly distributed in $[0,1]$, then compute $y = r_2 C$
3. If $y \leq f(x)$ then x is accepted, otherwise x is rejected and steps 1 and 2 are repeated until $y \leq f(x)$.

The rejection technique is always applicable, with any bounded pdf defined on a bounded interval $[a,b]$. However, if the pdf is strongly peaked, many pairs of random numbers might be generated before a successful trial.

Bibliography

[1] K. Tomizawa, *Numerical Simulation of Submicron Semiconductor Devices*. Boston: Artech House, 1993.

[2] C. Jacoboni, *Theory of Electron Transport in Semiconductors. A Pathway from Elementary Physics to Nonequilibrium Green Functions*. Berlin: Springer-Verlag, 2010.

[3] C. Hamaguchi, *Basic Semiconductor Physics*, 2nd ed. Berlin: Springer-Verlag, 2010.

[4] M. Fischetti and W. G. Vandenberghe, *Advanced Physics of Electron Transport in Semiconductors and Nanostructures*. Berlin: Springer-Verlag, 2016.

[5] C. Cercignani and E. Gabetta, Eds., *Transport Phenomena and Kinetic Theory: Applications to Gases, Semiconductors, Photons, and Biological Systems*. Boston: Birkhäuser, 2007.

[6] C. Jungemann and B. Meinerzhagen, *Hierarchical Device Simulation. The Monte-Carlo Perspective*, ser. Computational Microelectronics. Wien: Springer-Verlag, 2003.

[7] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, ser. Computational Microelectronics. Wien: Springer-Verlag, 1989.

[8] C. Moglestue, *Monte Carlo Simulation of Semiconductor Devices*. London: Chapman & Hall, 1993.

[9] F. Rossi, *Theory of Semiconductor Quantum Devices. Microscopic Modeling and Simulation Strategies*. Berlin: Springer-Verlag, 2011.

[10] L. Reggiani, Ed., *Hot-Electron Transport in Semiconductors*. Berlin: Springer-Verlag, 1985.

[11] R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem*, 2nd ed. New York: McGraw-Hill, 1976.

[12] F. Rossi and T. Kuhn, "Theory of ultrafast phenomena in photoexcited semiconductors," *Rev. Modern Phys.*, vol. 74, pp. 895–950, Aug. 2002.

[13] T. Kuhn, "Density matrix theory of coherent ultrafast dynamics," in *Theory of Transport Properties of Semiconductor Nanostructures*, E. Schöll, Ed. Dordrecht: Springer Science+Business Media, 1998, ch. 6, pp. 173–214.

[14] P. Brüesch, *Phonons: Theory and Experiments I. Lattice Dynamics and Models of Interatomic Forces*. Berlin: Springer-Verlag, 1982.

[15] B. C. Lee, K. W. Kim, M. Dutta, and M. A. Stroscio, "Electron–optical-phonon scattering in wurtzite crystals," *Phys. Rev. B*, vol. 56, no. 3, pp. 997–1000, Jul. 1997.

[16] I. Gorczyca, N. E. Christensen, E. L. Peltzer y Blancá, and C. O. Rodriguez, "Optical phonons modes in GaN and AlN," *Phys. Rev. B*, vol. 51, no. 17, pp. 11 936–11 939, May 1995.

[17] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors. Physics and Materials Properties*, 4th ed. Berlin: Springer-Verlag, 2010.

[18] G. L. Bir and G. E. Pikus, *Symmetry and Strain Induced Effects in Semiconductors*. New York: John Wiley & Sons, 1972.

[19] K. Hess, *Advanced Theory of Semiconductor Devices*. New York: IEEE, 2000.

[20] H. Haug and S. W. Koch, *Quantum theory of the optical and electronic properties of semiconductors*. Singapore: World Scientific Publishing, 2004.

[21] J.-J. Zhou and M. Bernardi, "Ab initio electron mobility and polar phonon scattering in GaAs," *Phys. Rev. B*, vol. 51, no. 94, p. 201201, 2016.

[22] M. V. Fischetti, P. D. Yoder, M. M. Khatami, G. Gaddemane, and M. L. Van de Put, "Hot electrons in Si lose energy mostly to optical phonons: Truth or myth?" *Appl. Phys. Lett.*, vol. 114, no. 6, p. 222104, 2019.

[23] J. Sjakste, N. Vast, and G. Barbarino, "Energy relaxation mechanism of hot-electron ensembles in GaAs: Theoretical and experimental study of its temperature dependence," *Phys. Rev. Lett.*, vol. 97, no. 6, p. 064302, 2018.

[24] B. K. Ridley, *Quantum Processes in Semiconductors*, 5th ed. Oxford: Oxford University Press, 2013.

[25] G. Ghione, *Dispositivi per la microelettronica*. Milano: McGraw-Hill, 1998.

[26] P. T. Landsberg, *Recombination in Semiconductors*. Cambridge, U.K.: Cambridge University Press, 1991.

[27] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors*. Berlin: Springer-Verlag, 1988.

[28] A. David, N. G. Young, C. Lund, and M. D. Craven, "Review–the physics of recombinations in III-nitride emitters," *ECS J. Solid State Sci. Technol.*, vol. 9, no. 1, p. 016021, 2020.

[29] J. Iveland, L. Martinelli, J. Peretti, J. S. Speck, and C. Weisbuch, "Direct measurement of Auger electrons emitted from a semiconductor light-emitting diode under electrical injection: Identification of the dominant mechanism for efficiency droop," *Phys. Rev. Lett.*, vol. 110, no. 17, p. 177406, Appl. Phys. Rev. 2013.

[30] M. Binder, A. Nirschl, R. Zeisel, T. Hager, H.-J. Lugauer, M. Sabathil, D. Bougeard, J. Wagner, and B. Galler, "Identification of *nnp* and *npp* Auger recombination as significant contributor to the efficiency droop in (GaIn)N quantum wells by visualization of hot carriers in photoluminescence," *Appl. Phys. Lett.*, vol. 103, no. 7, p. 071108, Aug. 2013.

[31] L. Meitner, "Über die entstehung der β-strahl-spektren radioaktiver substanzen," *Z. Phys.*, vol. 9, no. 1, pp. 131–144, 1922.

[32] P. Auger, "Sur les rayons secondaires produits dans un gaz par des rayons X," *C. R. Acad. Sci.*, vol. 177, pp. 169–171, 1923.

[33] D. B. Laks and G. F. Neumark, "Accurate interband-Auger-recombination rates in silicon," *Phys. Rev. B*, vol. 42, no. 8, pp. 5176–5185, Sep. 1990.

[34] C. Hamaguchi, *Basic Semiconductor Physics*. Berlin: Springer-Verlag, 2001.

[35] R. M. Yorston, "Free-flight time generation in the Monte Carlo simulation of carrier transport in semiconductors," *J. Comp. Phys.*, vol. 64, pp. 177–194, 1986.

[36] R. Fauquembergue, J. Zimmermann, A. Kaszynski, E. Constant, and G. Microondes, "Diffusion and the power spectral density and correlation function of velocity fluctuation for electrons in Si and GaAs by Monte Carlo methods," *J. Appl. Phys.*, vol. 51, no. 2, pp. 1065–1071, Feb. 1980.

[37] F. Bonani and G. Ghione, *Noise in Semiconductor Devices. Modeling and Simulation*, ser. Advanced Microelectronics. Berlin: Springer-Verlag, 2001.

[38] M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," *Phys. Rev. B*, vol. 38, no. 14, pp. 9721–9745, Nov. 1988.

[39] M. V. Fischetti, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures—Part I: Homogeneous transport," *IEEE Trans. Electron Devices*, vol. 38, no. 3, pp. 634–649, Mar. 1991.

[40] T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, and C. Hamaguchi, "A Monte Carlo simulation of anisotropic electron transport in silicon including full band structure and anisotropic impact-ionization model," *J. Appl. Phys.*, vol. 75, no. 1, pp. 297–312, Jan. 1994.

[41] D. Esseni and P. Palestri, "Theory of the motion at the band crossing points in bulk semiconductor crystals and in inversion layers," *J. Appl. Phys.*, vol. 105, no. 5, p. 053702, 2009.

[42] H. Shichijo and K. Hess, "Band-structure-dependent transport and impact ionization in GaAs," *Phys. Rev. B*, vol. 23, no. 8, pp. 4197–4207, 1981.

[43] G. Ghione, *Semiconductor Devices for High-Speed Optoelectronics*. Cambridge, U.K.: Cambridge University Press, 2009.

[44] R. McIntyre, "A new look at impact ionization - Part I: a theory of gain, noise, breakdown probability, and frequency response." *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1623–1631, Aug. 1999.

[45] M. Hayat, B. Saleh, and M. Teich, "Effect of dead space on gain and noise of double-carrier-multiplication avalanche photodiode," *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 546–551, May 1992.