TECHNISCHE UNIVERSITÄT MÜNCHEN
Physik-Department
Institut für Theoretische Physik
Lehrstuhl Univ.-Prof. Dr. Peter Vogl

Modeling of semiconductor nanostructures and semiconductor–electrolyte interfaces

Stefan Birner

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

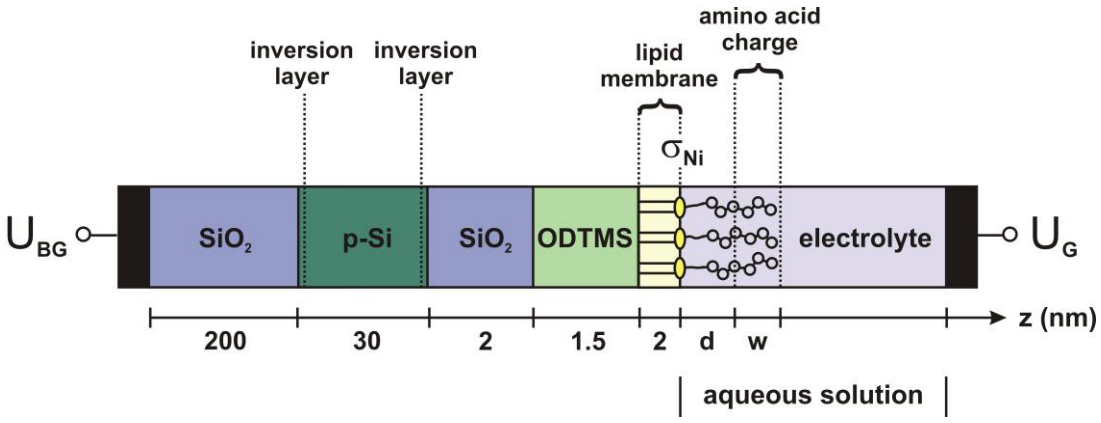
genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Jonathan J. Finley, Ph.D.
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Peter Vogl 2. Univ.-Prof. Paolo Lugli, Ph.D.

Die Dissertation wurde am 12.10.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 15.11.2011 angenommen.

Modeling of semiconductor nanostructures and semiconductor–electrolyte interfaces

Stefan Birner



Cover picture

Schematic layout of a silicon-on-insulator based protein sensor.

Two inversion layers are present in the *p*-type silicon layer. The density of these two-dimensional electron gases (2DEGs) can be tuned by adjusting the back gate U_{BG} and the electrolyte gate potential U_G. Ni^{2+} ions in the electrolyte solution form a chelate complex with the NTA headgroups of the functionalized surface resulting in a negative interface charge density σ_{Ni} at the lipid–electrolyte interface. The amino acid charge is assumed to be distributed homogeneously over a width *w*. The electrolyte region includes the histidine-tagged amino acids as well as the neutral part of the tag of length *d*. Upon adsorption of amino acids the density in the right inversion layer changes slightly which increases the conductivity, and thus the source–drain current in the 2DEG.

Contents

Abstract

This thesis consists of two parts.

The main objective of Part I is to give an overview of some of the methods that have been implemented into the next**nano**^{3} software. Examples are discussed that give insight into doping, strain and mobility. Applications of the single-band Schrödinger equation include three-dimensional superlattices, and a qubit that is manipulated by a magnetic field. Results of the multi-band $\mathbf{k} \cdot \mathbf{p}$ method are presented for HgTe–CdTe and InAs–GaSb superlattices, and for a SiGe–Si quantum cascade structure. Particular focus is put on a detailed description of the contact block reduction (CBR) method that has been developed within our research group. By means of this approach, quantum transport in the ballistic limit in one, two and three dimensions can be calculated. I provide a very detailed description of the algorithm and present several well documented examples that highlight the key points of this method. Calculating quantum transport in three dimensions is a very challenging task where computationally efficient algorithms – apart from the CBR method – are not available yet.

Part II describes the methods that I have implemented into the next**nano**^{3} software for calculating systems that consist of a combination of semiconductor materials and liquids. These biosensors have a solid–electrolyte interface, and the charges in the solid and in the electrolyte are coupled to each other through the Poisson–Boltzmann equation. I apply this model to a silicon based protein sensor, where I solve the Schrödinger equation together with the Poisson–Boltzmann equation self-consistently, and compare theoretical results with experiment. Furthermore, I have developed a novel approach to model the charge density profiles at semiconductor–electrolyte interfaces that allows us to distinguish hydrophobic and hydrophilic interfaces. Our approach extends previous work where ion specific potentials of mean force describe the distribution of ion species at the interface. I apply this new model to recently developed graphene and diamond based solution gated field-effect transistors, and compare my calculations to experiment.

All numerical examples presented in this thesis are available as input files for the next**nano**^{3} and/or next**nano**++ software. It is thus possible for other researchers to reproduce the results of all calculations of this thesis. Additionally, the respective input files can easily be modified to study variations of device characteristics, like geometry, choice of materials, doping, and many more. To date, the next**nano** software has been used successfully in many master and doctoral theses, as well as in numerous scientific articles to provide either a qualitative understanding or a quantitative analysis of the electronic and optoelectronic properties of modern semiconductor nanostructures.

Zusammenfassung

Diese Arbeit besteht aus zwei Teilen.

In Teil I stelle ich einige Modelle vor, die in der next**nano**-Software verwendet werden. Als Beispiele diskutiere ich dotierte und verspannte Halbleiter, sowie die Mobilität von Ladungsträgern. Dreidimensionale Übergitter und ein durch ein Magnetfeld manipuliertes Qubit werden als Anwendungen der Einband-Schrödinger-Gleichung präsentiert. Mit Hilfe der Mehrband-$\mathbf{k} \cdot \mathbf{p}$-Schrödinger-Gleichung berechne ich HgTe–CdTe- und InAs–GaSb-Übergitter, sowie eine SiGe–Si-Quantenkaskadenstruktur. Besonderes Augenmerk lege ich auf die detaillierte Beschreibung der 'Contact Block Reduction'-Methode (CBR), die in unserer Arbeitsgruppe entwickelt wurde. Sie erlaubt, ballistischen Quantentransport in ein, zwei und sogar drei Dimensionen zu berechnen. Der entsprechende Algorithmus wird sorgfältig anhand mehrerer Beispiele erklärt, um die wesentlichen Aspekte hervorzuheben. Quantentransportrechnungen in drei Dimensionen sind extrem aufwendig und mit momentan verfügbaren Rechenkapazitäten kaum zu meistern. Außer der CBR-Methode sind derzeit keine effizienten Algorithmen verfügbar.

Teil II behandelt die Modelle, die ich in next**nano** implementiert habe, um Systeme zu berechnen, die aus einer Kombination von Halbleitermaterialien und Flüssigkeiten bestehen. Diese Biosensoren weisen eine Festkörper–Elektrolytgrenzfläche auf, wobei die Ladungen im Festkörper und in der Flüssigkeit über die Poisson–Boltzmann-Gleichung gekoppelt sind, die zusammen mit der Schrödinger-Gleichung selbstkonsistent gelöst wird. Als Anwendung berechne ich siliziumbasierte Proteinsensoren. Des Weiteren habe ich eine neue Methode entwickelt, um die Verteilung der Ladungsträger an Halbleiter–Elektrolytgrenzflächen zu modellieren. Sie erlaubt es, zwischen hydrophoben und hydrophilen Grenzflächen zu unterscheiden. Mittels ionenspezifischer Potenziale des mittleren Feldes wird die Verteilung der Ionen an der Grenzfläche beschrieben. Damit analysiere ich Biosensoren, die aus Graphen und Diamant bestehen, und vergleiche die theoretischen Ergebnisse mit experimentellen Resultaten.

Die in dieser Arbeit berechneten Beispiele sind als Eingabedateien für die next**nano**-Software verfügbar. Somit können alle hier vorgestellten Ergebnisse reproduziert werden. Die entsprechenden Eingabedateien können modifiziert werden, um Variationen der Bauelementeigenschaften, wie z. B. Geometrie, Wahl der Materialien, Dotierungen, usw. zu untersuchen. next**nano** wurde bereits in zahlreichen Diplom- und Doktorarbeiten, sowie wissenschaftlichen Veröffentlichungen erfolgreich verwendet, um sowohl ein qualitatives Verständnis, als auch eine quantitative Analyse der elektronischen und optoelektronischen Eigenschaften moderner Halbleiternanostrukturen zu ermöglichen.

Introduction

The quickly progressing technology of semiconductor quantum structures requires and depends on reliable predictive theoretical methods for systematically improving, designing and understanding the electronic and optical properties of such structures. Due to the increase in computing power and the simultaneous decrease of cost for fast processors and memory, computational physics is no longer a field of specialists (i.e. theorists), that have access to powerful supercomputers. Computer modeling nowadays has become a convenient tool for both, educational purposes as well as to support experimentalists while analyzing measured data or to design new experiments. The challenge is to make available to this audience a tool that covers the most important semiconductor equations, including the related theoretical improvements that have been made over the past decades, for instance the ones that go beyond the simple 'single-band effective-mass' model that is still widely used due to its simplicity and intuitiveness[1]. These more sophisticated models take into account the anisotropy and nonparabolicity of the electron and hole masses, usually employed within an 8-band $\mathbf{k} \cdot \mathbf{p}$ model. Strain is an important degree of freedom to optimize the electronic (e.g. mobility) and optical properties (e.g. transition energies) of heterostructures. This involves a model to take into account deformation potentials as well as piezoelectric fields. The technologically important nitride materials crystallize in the wurtzite structure. This introduces more complexity into the equations in comparison to the simpler zinc blende materials as the cubic symmetry is lost. Additionally, pyroelectric fields have to be taken into account for these materials. Nowadays, semiconductor layers can be grown not only on (001) oriented wafers but also along less symmetric crystallographic directions like [110] or [311]. These orientations not only require the rotation of the coordinate systems and the relevant equations, they also make it necessary to consider nonsymmetric displacement tensors as well as piezoelectric fields. The variety of possible semiconductor materials and their alloys (including organic semiconductors) for designing and improving devices is huge. A software has to take care of all of them, and should provide an up-to-date database with all relevant material parameters. This is indeed very relevant for the topic of 'sustainable materials', which includes the reduction and eventually elimination of the use of toxic materials, like the replacement of CdS by ZnS in CIGS solar cells, or the replacement of rare materials like indium with abundant alternatives. Such a replacement will inevitably lead to a decrease in performance which requires a detailed understanding of the device characteristics in order to again enhance performance by modifying the device layout. Certainly, the detailed analysis of such structures by means of modeling is an excellent way to support such activities.

[1] e.g. G. Snider's '1D Poisson' program. It can be obtained from: http://www.nd.edu/~gsnider/

In addition to one-dimensional quantum confinement (e.g. quantum wells, superlattices), two-dimensional and three-dimensional quantum confinement has been studied intensively during the past decades both experimentally and theoretically (quantum wires, quantum dots, quantum dot crystals). Consequently, the latter heterostructures require 2D or 3D simulation environments (e.g. [SGB99]) which go far beyond simple self-written 1D codes for quantum wells. Most of these structures require the application of a bias, and thus a model that calculates the current has to be implemented.

In this thesis a selection of the relevant physical models that are implemented in the software package next**nano** are described. next**nano** allows one to study the realistic electronic structure and optical properties of arbitrarily shaped three-dimensional semiconductor nanostructures consisting of diamond-type, zinc blende or wurtzite materials and their alloys. It is possible to study these heterostructures under bias and calculate the current density close to equilibrium. First, the strain is calculated within a continuum elasticity approach. Using band offsets and deformation potential theory, the new conduction and valence band edges are obtained. Then the Poisson, multi-band Schrödinger and current equations are solved self-consistently, taking into account doping, piezo- and pyroelectric charges, excitonic effects and magnetic fields. Finally, optical properties such as transition matrix elements or absorption can be calculated. Describing the electronic transport on a quantum mechanical level that takes into account scattering on an equal footing – in contrast to ballistic transport – is still a challenge. The quantum cascade laser is a nice example of a quantum device that has been invented only recently. It is for sure that further devices will be engineered that profoundly rely on quantum mechanics. They are expected to path the way to new disruptive technologies. Possible candidates might include the fields of quantum information processing, spintronics, thermoelectrics or nanostructures for energy generation, conversion or storage. Structures like biosensors that connect the semiconductor world to the life sciences are expected to have a promising future and economical potential, but developing these devices requires a very interdisciplinary approach. A software for biosensors that works as an educational tool could provide the chemistry and life science experts with a decent introduction into the field of semiconductor physics, thus contributing to make the designing process of such devices more efficient. Theoretical modeling of solid-state qubits – the building blocks of quantum computers – is very likely of great help to experimentalists, whereas actually building a quantum computer still remains one of the greatest challenges of current and future research activities in physics. The next**nano** software has been, is and will be of real benefit to the semiconductor community. This thesis highlights some of the models implemented, and discusses recent applications and results.

Part I.

Semiconductors

1. Calculation of semiconductor nanostructures

In this chapter we sketch the capabilities of the next**nano** software and discuss some of the main equations that are implemented into the code. We illustrate them using several examples, like the charge carrier concentrations for different doping properties, or the strain tensor components of a compressively strained layer for different growth directions.

next**nano** is a simulation tool that aims at providing global insight into the basic physical properties of realistic three-dimensional mesoscopic semiconductor structures. It focuses on quantum mechanical properties such as the global electronic structure, optical properties [Eiß08], and the effects of electric and magnetic fields for virtually any geometry and combination of semiconducting materials. For the calculation of the carrier dynamics a drift–diffusion model based on a quantum-mechanically calculated density is employed [Hac02]. Alternatively, a self-consistent ballistic transport model based on the contact block reduction (CBR) method can be used (Chapter 4). A detailed discussion of the implemented models were already presented in previous PhD theses by Hackenbuchner [Hac02], Sabathil [Sab04], Zibold [Zib07] and Andlauer [And09]. For that reason we chose to avoid reproducing their equations and focus on aspects of the next**nano** software not previously covered, like a summary of all the various sets of $\mathbf{k} \cdot \mathbf{p}$ parameters used in the literature, examples for strain calculations for arbitrary substrate orientations, examples for doping, calculations of minibands in superlattices, self-consistent $\mathbf{k} \cdot \mathbf{p}$ Schrödinger–Poisson calculations for several substrate orientations, comparison of $\mathbf{k} \cdot \mathbf{p}$ calculations to tight-binding calculations, and finally the modeling of semiconductor–electrolyte structures (Part II).

Several empirical band structure methods for semiconductors exist. The three most common ones are the tight-binding, the pseudopotential and the $\mathbf{k} \cdot \mathbf{p}$ method. The next**nano** software and this thesis are based on the (multi-band) $\mathbf{k} \cdot \mathbf{p}$ method where the basis functions used to represent the Schrödinger equation are Bloch states (Chapter 3). Pseudopotential calculations are based on plane-wave basis functions whereas tight-binding uses atom-like basis functions. The wave functions obtained from the $\mathbf{k} \cdot \mathbf{p}$ Schrödinger equation are envelope functions. Therefore, this method is also known as the multi-band envelope function approximation (EFA). In the special case of taking into account only one band (usually a conduction band), the multi-band $\mathbf{k} \cdot \mathbf{p}$ method reduces to the single-band Schrödinger equation (effective-mass EFA) (Chapter 2).

Once the number of atoms of a device exceeds a few hundreds, microscopic methods that take into account the detailed atomic structure become not only unfeasible altogether but also impractical as the observables of interest are slowly varying over atomic

distances. In spite of many attempts to develop alternatives, there is still only one established and well-studied electronic structure method that is suitable for systems of mesoscopic length scales. This approximation widely used for the past 50 years is the so-called envelope function approximation based on the $\mathbf{k} \cdot \mathbf{p}$ method which is described in more detail in Chapter 3. In spite of its known limitations and shortcomings [WZ96], this method has been amazingly successful in predicting semiconductor nanostructures (e.g. [Bas88]). The EFA turned out to be a powerful tool for device simulations since usually we are not interested in the details of inter-atomic distance scale. The most attractive feature of the EFA is relatively low computationally cost both in terms of CPU time and memory size.

1.1. Poisson equation

The Poisson equation describes the electrostatics within the device and reads

$$\nabla \cdot [\varepsilon_0 \varepsilon_\mathrm{r}(\mathbf{x}) \nabla \phi(\mathbf{x})] = -\rho(\mathbf{x}), \tag{1.1}$$

where ϕ is the electrostatic potential and ε_0 is the vacuum permittivity. The tensor ε_r is the material dependent static dielectric constant at position $\mathbf{x}$. It is isotropic in zinc blende materials but anisotropic for wurtzite.

The charge density distribution $\rho(\mathbf{x})$ within a semiconductor device is given by

$$\rho(\mathbf{x}) = e \left[-n(\mathbf{x}) + p(\mathbf{x}) + N_\mathrm{D}^+(\mathbf{x}) - N_\mathrm{A}^-(\mathbf{x}) + \rho_\mathrm{fix}(\mathbf{x}) \right], \tag{1.2}$$

where e is the positive elementary charge, n and p are the electron and hole densities, and N_D^+ and N_A^- are the ionized donor and acceptor concentrations, respectively. The latter are covered in the next section. If required, fixed interface or volume charge densities ρ_fix can be taken into account, e.g. arising from piezo- or pyroelectric charges. The electron and hole densities can be calculated classically within the Thomas–Fermi approximation or quantum mechanically if quantum confinement effects are important [Hac02].

We discretize the Poisson equation on a nonuniform grid with a finite differences method and solve it numerically by an iterative Newton–Raphson scheme. More details about the numerical solution of the Poisson equation can be found in Refs. [Hac02, LKBJ97, TSCH90]. Typically, Neumann boundary conditions for the Poisson equation are employed which implies a vanishing electric field at the boundaries

$$\frac{\partial \phi}{\partial x} = 0. \tag{1.3}$$

For nonequilibrium simulations we use Dirichlet boundary conditions [LKBJ97]. Here, one first has to determine the electrostatic potential in equilibrium (built-in potential) using zero-field (Neumann) boundary conditions. The electrostatic potential at the boundaries is then fixed (Dirichlet boundary condition) with respect to the chemical potentials taking into account the previously calculated built-in potential at the boundaries [Hac02]. The chemical potentials at the contacts are fixed and correspond to the applied bias. An example of a Schottky barrier boundary condition is discussed and simulated in Appendix A.

1.2. Doping

The ionized shallow donor and acceptor densities N_D^+, N_A^- are given by

$$N_\mathrm{D}^+(\mathbf{x}) = \sum_{i \in \mathrm{Donors}} \frac{N_{\mathrm{D},i}(\mathbf{x})}{1 + g_{\mathrm{D},i} \exp\left(\left(E_\mathrm{F,n}(\mathbf{x}) - E_{\mathrm{D},i}(\mathbf{x})\right)/k_\mathrm{B}T\right)} \tag{1.4}$$

$$N_\mathrm{A}^-(\mathbf{x}) = \sum_{i \in \mathrm{Acceptors}} \frac{N_{\mathrm{A},i}(\mathbf{x})}{1 + g_{\mathrm{A},i} \exp\left(\left(E_{\mathrm{A},i}(\mathbf{x}) - E_\mathrm{F,p}(\mathbf{x})\right)/k_\mathrm{B}T\right)}, \tag{1.5}$$

where the summation is over all different donor or acceptors species i, k_B is the Boltzmann constant, T is the temperature, $E_\mathrm{F,n}$, $E_\mathrm{F,p}$ are the electron and hole Fermi levels, $N_{\mathrm{D},i}$, $N_{\mathrm{A},i}$ are the donor and acceptor concentrations and $E_{\mathrm{D},i}$, $E_{\mathrm{A},i}$ are the energies of the neutral donor and acceptor impurities, respectively, that generally also depend on $\mathbf{x}$. The latter are determined by the ionization energies $E_{\mathrm{D},i}^\mathrm{ion}$, $E_{\mathrm{A},i}^\mathrm{ion}$, the bulk conduction and valence band edges (including shifts due to strain) and the electrostatic potential. The ionization energies of these shallow donors $E_{\mathrm{D},i}^\mathrm{ion}$, $E_{\mathrm{A},i}^\mathrm{ion}$ are rather small (between $5 - 50\,\mathrm{meV}$) and thus one can assume that the impurities are in equilibrium with the conduction or valence bands that are energetically very close. Therefore the number of ionized donors depends on the local quasi-Fermi level of the electrons, and the number of ionized acceptors on the local quasi-Fermi level of the holes

$$E_{\mathrm{D},i}(\mathbf{x}) = E_{\mathrm{c},0}(\mathbf{x}) - e\phi(\mathbf{x}) - E_{\mathrm{D},i}^\mathrm{ion}(\mathbf{x}) = E_\mathrm{c}(\mathbf{x}) - E_{\mathrm{D},i}^\mathrm{ion}(\mathbf{x}) \tag{1.6}$$

$$E_{\mathrm{A},i}(\mathbf{x}) = E_{\mathrm{v},0}(\mathbf{x}) - e\phi(\mathbf{x}) + E_{\mathrm{A},i}^\mathrm{ion}(\mathbf{x}) = E_\mathrm{v}(\mathbf{x}) + E_{\mathrm{A},i}^\mathrm{ion}(\mathbf{x}). \tag{1.7}$$

The degeneracy factor is usually given by $g_{\mathrm{D},i} = 2$ for donors and $g_{\mathrm{A},i} = 4$ for acceptors [Böe92] and considers the different transition rates from the impurity into the conduction or valence band and vice versa.
Shallow donors ($g_{\mathrm{D},i} = 2$): The outer s orbital is onefold occupied in the neutral state. There is one possibility to get rid of one electron but there are two to incorporate one electron (spin up $\uparrow$, spin down $\downarrow$).
Shallow acceptors ($g_{\mathrm{A},i} = 4$): The sp^3 orbital is threefold occupied. Thus there is one possibility to incorporate an electron and four possibilities to get rid of an electron.

We assume that both donor and acceptor impurities have only a single energy level $E_{\mathrm{D},i}$, $E_{\mathrm{A},i}$ degenerated by factors $g_{\mathrm{D},i}$, $g_{\mathrm{A},i}$. For the general case, an impurity can have several different energy levels. In nitride semiconductors crystallizing in the wurtzite structure, g_A may vary from 4 to 6 due to a small valence band splitting [Sem04].

Now we present results of our calculations on the temperature dependence of the carrier concentrations in doped bulk germanium samples. Figure 1.1 shows the electron n and hole concentrations p as a function of temperature. (The melting point of Ge is at 1211.40 K.) The Ge sample is n-type doped with a concentration of $N_\mathrm{D} = 1 \cdot 10^{17}\,\mathrm{cm}^{-3}$. The ionization energy of the donor is set to $E_\mathrm{D}^\mathrm{ion} = 10\,\mathrm{meV}$ as in Ref. [Gra99]. In the saturation regime between 150 K and 600 K, the charge carrier density is equal to the doping density, i.e. $n \cong 1 \cdot 10^{17}\,\mathrm{cm}^{-3}$, i.e. almost all donors are ionized. For large temperatures the intrinsic carrier density dominates, i.e. here the electron density is equal

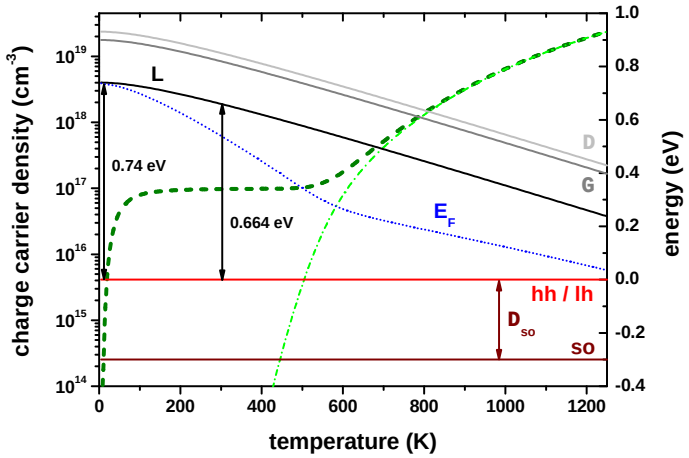


Figure 1.1.: Electron (dashed line) and hole (dash-dotted) concentrations in n-type doped Ge as a function of temperature. The band edges (solid lines) reveal the temperature dependence of the band gap. For low temperatures, the position of the Fermi level E_F (dotted line) is close to the L conduction band edge. At higher temperatures it is around the middle of the band gap.

to the hole density. The figure also shows the band gaps of the Γ, L and Δ conduction band minima vs. temperature which decrease with increasing temperature. The valence band edges are also shown: Heavy (hh) and light hole (lh) are degenerate, separated by Δ_so from the split-off (so) hole band. The temperature dependence of the band gap is described using the Varshni formula (eq. (B.2)). For very low temperatures, the position of the Fermi level E_F (chemical potential, blue dotted line), $E_\mathrm{F}(T = 0\,\mathrm{K}) = E_\mathrm{c}^\mathrm{L} - E_\mathrm{D}^\mathrm{ion}/2$, is close to the L conduction band edge, i.e. it is located between the donor level and the conduction band edge. For temperatures between 600 K and 900 K, the Fermi level is around the middle of the band gap.

An interesting situation occurs in real semiconductors that have both, n-type and p-type doping. It is rather impossible to fabricate devices that are merely n- or p-type because unintentional doping through impurities is always present. These semiconductors are called *compensated* semiconductors. We assume the same Ge sample as above, i.e. the Ge sample is n-type doped with the same concentration of $N_\mathrm{D} = 1 \cdot 10^{17}\,\mathrm{cm}^{-3}$ and the same ionization energy. Now we also include acceptors with varying acceptor concentration ranging from $N_\mathrm{A} \ll N_\mathrm{D}$ and $N_\mathrm{A} < N_\mathrm{D}$ to $N_\mathrm{A} = N_\mathrm{D}$, i.e. we now have one donor level and one acceptor level. The ionization energy of the acceptor is set to $E_\mathrm{A}^\mathrm{ion} = 10\,\mathrm{meV}$. Figure 1.2 shows the calculated electron density vs. temperature for different acceptor concentrations N_A. If $N_\mathrm{A} = N_\mathrm{D}$, then the intrinsic behavior of the electron density is recovered (fully compensated). In the saturation regime, the nominal dopant density $N_\mathrm{D} - N_\mathrm{A}$ determines the electron density (partially compensated).

Now we examine a compensated Ge sample with two donor levels and one acceptor

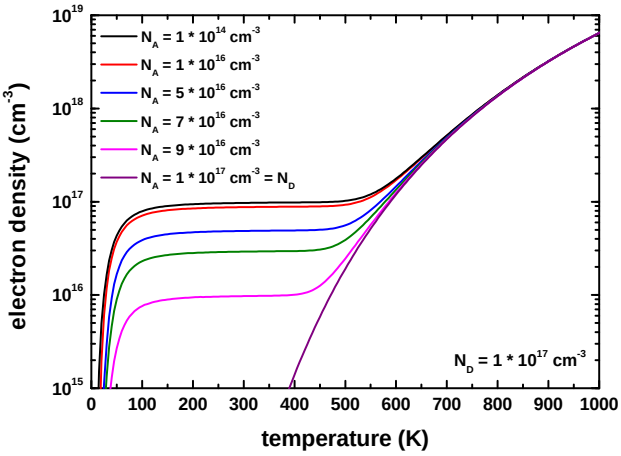


Figure 1.2.: Compensated semiconductors: Electron density vs. temperature for different acceptor concentrations N_A and constant donor concentration N_D.

level. The acceptor concentration is fixed to $N_\mathrm{A} = 5 \cdot 10^{16}\,\mathrm{cm}^{-3}$ but the donor concentrations $N_{\mathrm{D},1}$ and $N_{\mathrm{D},2}$ are varied. The ionization energy of the acceptor is set to $E_\mathrm{A}^\mathrm{ion} = 10\,\mathrm{meV}$. The ionization energy of the upper donor level is $E_{\mathrm{D},1}^\mathrm{ion} = 10\,\mathrm{meV}$, the one of the lower donor level $E_{\mathrm{D},2}^\mathrm{ion} = 100\,\mathrm{meV}$. We distinguish three cases. The case where $N_{\mathrm{D},2} = 0$ (black solid line) is similar to the figures that have been discussed already. Its saturation regime has a dopant density of $N_{\mathrm{D},1} - N_\mathrm{A} = 6 \cdot 10^{16}\,\mathrm{cm}^{-3}$. For the second case (blue solid line) where the concentration of the upper donor level is smaller than the concentration of the acceptor level, only the ionization of the second donor level $E_{\mathrm{D},2}$ is relevant at low temperatures. For the case where both donor levels have a concentration higher than the acceptor concentration, two plateaus form. The first is determined by the ionization of the upper donor level and leads to a donor density of $N_{\mathrm{D},1} - N_\mathrm{A} = 1 \cdot 10^{16}\,\mathrm{cm}^{-3}$. The second plateau is determined by the ionization of the lower donor level. Its donor density is $N_{\mathrm{D},1} + N_{\mathrm{D},2} - N_\mathrm{A} = 6 \cdot 10^{16}\,\mathrm{cm}^{-3}$. For comparison, for the last case the energetic position of the second donor level $E_{\mathrm{D},2}$ has been varied between 70 meV (red dotted line), 100 meV (red solid line) and 130 meV (red dashed line). The energetic position $E_{\mathrm{D},2}$ of the upper donor level is always at 10 meV.

Figure 1.4 shows the calculated position of the Fermi level with respect to the band gap for the cases discussed above. The valence band edge is fixed at 0 eV as in Fig. 1.1. For the blue solid line where at low temperatures only the ionization of the second donor level $E_{\mathrm{D},2}$ is relevant, one can clearly see the position of the Fermi level, $E_\mathrm{F}(T = 0\,\mathrm{K}) = E_\mathrm{c} - E_{\mathrm{D},2}^\mathrm{ion}$, at 100 meV below the conduction band edge. For the red curves, the Fermi level moves from the upper donor level $E_{\mathrm{D},1}$ at low temperatures to the lower donor level $E_{\mathrm{D},2}$ at around 200 K and starts to ionize the second donor level, eventually forming the second plateau in the electron density (see Fig. 1.3). At high temperatures, the Fermi levels moves approximately into the middle of the band gap.

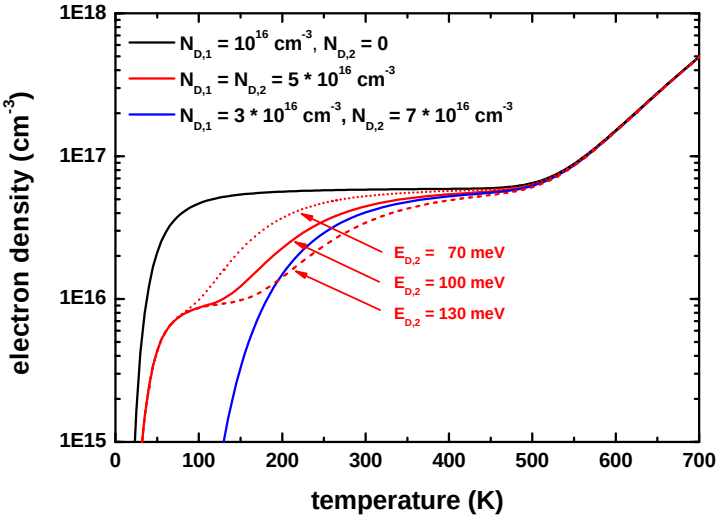


Figure 1.3.: Calculated electron density vs. temperature for two donor levels and one acceptor level.

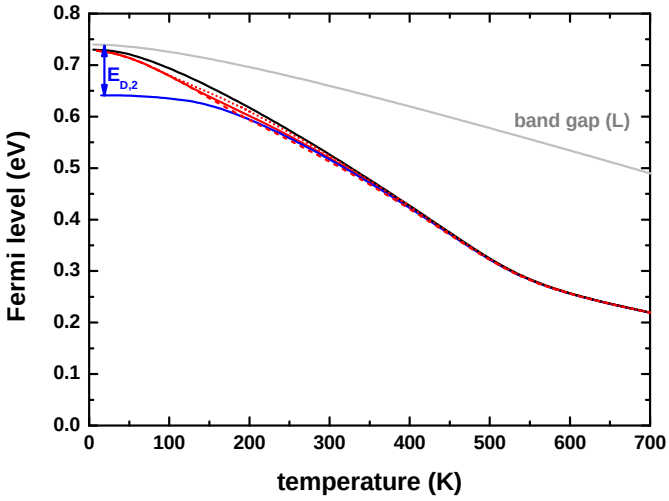


Figure 1.4.: Calculated position of the Fermi level with respect to the band gap vs. temperature for two donor levels and one acceptor level.

Obviously, for realistic devices with complicated doping profiles where an intuitive understanding of charge redistribution is lacking, it is very convenient to have a tool by hand – like the next**nano** software – that can handle eq. (1.4) and eq. (1.5) numerically.

1.3. Strain equation

Strain can be used to alter and optimize the electronic and optical properties of quantum wells, wires and dots by varying both the energy levels and the spatial extent of the wave functions. Heterostructures made of semiconductor materials with different lattice constants are subject to elastic deformations. Such deformations can be studied within a continuum mechanical model based on classical elasticity within the harmonic approximation, i.e. for small strains. The lattice deformation changes drastically with the surface orientation (i.e. growth direction). An exact knowledge of the strain fields is of great importance for the interpretation of experimental data and for the design and development of optoelectronic and electronic devices. Strain leads to piezoelectric effects, influences the conduction and valence band edges (including their degeneracies) [VdW89] and the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian of Schrödinger's equation [Hac02]. Therefore strain is a very important 'tool' for device engineers to alter the electronic (e.g. mobility, see Chapter 5) and optical properties of semiconductor heterostructures. The next**nano** software calculates the strain prior to the Poisson, Schrödinger or current equations, i.e. strain is independent of all other equations and can thus be separated from the main part of the program. This separation might not be fully justified for wurtzite structures where strong pyroelectric fields exist. In this case, the strain and Poisson equations have to be solved self-consistently, as demonstrated by Willatzen et al. [WLLYVM06], leading to corrections to the energy levels of the order of several meV in nitride semiconductor structures.

The components of the strain tensor $\epsilon(\mathbf{x})$, i.e. the symmetrical part of the distortion tensor $\tilde{\mathbf{u}}(\mathbf{x})$ (eq. (C.18)), are defined as

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) = \frac{1}{2}\left(\tilde{u}_{ij} + \tilde{u}_{ji}\right) = \epsilon_{ji}, \tag{1.8}$$

where $i, j = 1, 2, 3$. The vector $\mathbf{u}(\mathbf{x})$ describes the displacement due to lattice deformations. The strain tensor ϵ is symmetric, whereas the distortion tensor $\tilde{\mathbf{u}}$ is in general not symmetric. The latter is, however, identical to the strain tensor for high symmetry substrate orientations like (001), (110) and (111). It is not identical for any of the other low symmetry orientations. The diagonal elements of the strain tensor measure the extensions per unit length along the coordinate axes (positive values mean tensile strain, negative values compressive strain), i.e. the lengths of the considered volume element change while the angles remain constant. In contrast, the off-diagonal elements measure the shear deformations where the angles change and the volume remains constant. As the 3×3 matrix ϵ is real and symmetric, it can always be diagonalized by an appropriate orthogonal transformation. The resulting diagonal matrix must not necessarily coincide with the crystal or simulation coordinate system (compare with Appendix C). The trace of the strain tensor $\mathrm{Tr}(\epsilon)$ is equal to the hydrostatic strain, i.e. the change in volume. The strain tensor components are obtained by minimizing the elastic energy E

$$E = \frac{1}{2}\int_V C_{ijkl}\epsilon_{ij}\epsilon_{kl}\,\mathrm{d}V, \tag{1.9}$$

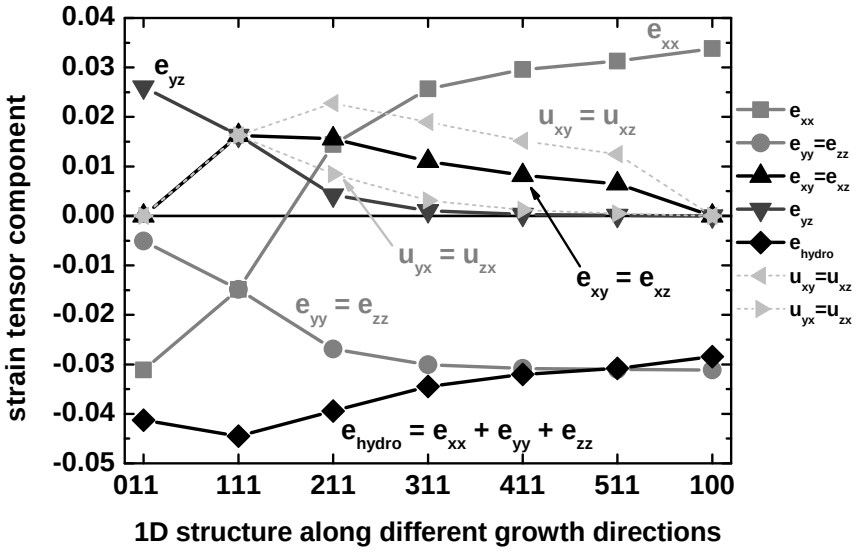


Figure 1.5.: Strain tensor components with respect to the crystal coordinate system for a compressively strained InAs layer grown pseudomorphically on InP for different growth directions. Note that [100] growth direction corresponds to [∞11] and only for this growth direction all off-diagonal strain tensor components are zero. The maximum value for the volume deformation (hydrostatic strain) is obtained for [111]. (The lines are a guide to the eye.)

where C_{ijkl} is the forth-ranked elastic stiffness tensor. A derivation of the numerical calculation of the strain tensor for arbitrary three-dimensional heterostructures can be found in Ref. [Hac02].

Analytical equations of the strain and distortion tensors with respect to the *crystal* coordinate system are given for cubic crystals in Appendix C.2. Additionally, I derived analogous equations with respect to the *simulation* coordinate system [PGBD$^+$11] (Appendix C.1). These equations are valid for heterostructures that are homogeneous along two directions, and for arbitrary substrate orientations.

We now present results for a compressively strained InAs layer grown pseudomorphically on InP for different growth directions. We plot the strain tensor components with respect to the crystal coordinate system, and with respect to the simulation coordinate system. The first is identical to the conventional cubic cell basis. The simulation coordinate system is chosen so that the growth direction of the heterostructure is perpendicular to the plane of the substrate.

Figure 1.5 shows the strain tensor components with respect to the *crystal* coordinate system. The maximum value for the volume deformation (i.e. extremal value of the hydrostatic strain) is obtained for the [111] growth direction. In zinc blende heterostruc-

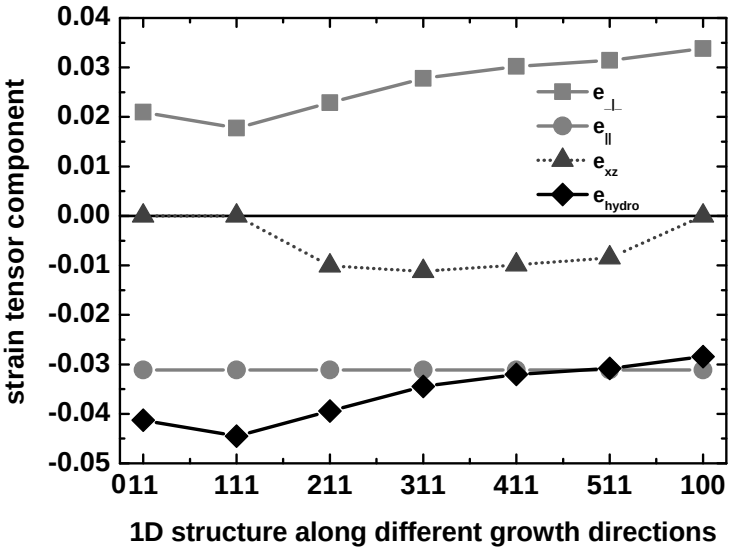


Figure 1.6.: Strain tensor components for the same situation as in Fig. 1.5 but this time with respect to the simulation coordinate system. For the high symmetry substrate orientations [100], [011] and [111], all off-diagonal strain tensor components are zero. (The lines are a guide to the eye.)

tures grown along the [100] direction only diagonal strain components occur but any other growth direction exhibits off-diagonal strain components that lead to piezoelectric polarization.

Figure 1.6 shows the strain tensor components with respect to the *simulation* coordinate system for the same structure. Here, the x axis is assumed to be the growth direction. The hydrostatic strain must of course coincide with Fig. 1.5 and its maximum value for the volume deformation is obtained for [111] (as before), where the perpendicular strain component $\epsilon_\perp$ has its minimum. For the high symmetry substrate orientations [100], [011] and [111], all off-diagonal strain tensor components are zero. The strain tensor components in the plane parallel to the substrate plane are independent of growth direction for cubic crystals and are given by the lattice mismatch

$$\epsilon_\parallel = \frac{a_{\text{substrate}} - a_{\text{layer}}}{a_{\text{layer}}}, \tag{1.10}$$

where $a_{\text{substrate}}$ and a_{layer} are the lattice constants of the substrate and of the strained layer material. For substrate orientations that have even lower symmetry, like e.g. (321), also the off-diagonal component ϵ_{xy} would be nonzero. The off-diagonal component ϵ_{yz} is always zero (see Appendix C.1).

2. Single-band envelope function approximation

2.1. The single-band Schrödinger equation

The single-band Schrödinger equation is a special case of the 8×8 $\mathbf{k} \cdot \mathbf{p}$ Schrödinger equation that is described in the next chapter. If the coupling between the conduction and valence bands is ignored, one obtains a twofold (due to spin) degenerate Schrödinger equation for the electrons in a heterostructure

$$\mathbf{H}\Psi_n(\mathbf{x}) = E_n\Psi_n(\mathbf{x}) \tag{2.1}$$

$$\left(-\frac{\hbar^2}{2}\nabla \cdot \mathbf{M}(\mathbf{x})\nabla + V(\mathbf{x})\right)\Psi_n(\mathbf{x}) = E_n\Psi_n(\mathbf{x}). \tag{2.2}$$

The potential energy $V(\mathbf{x})$ is the resulting conduction band edge profile that includes band offsets, band shifts due to strain, and the electrostatic potential. The effective mass is described by the tensor $\mathbf{M}$

$$\mathbf{M} = \begin{pmatrix} 1/m_{xx} & 1/m_{xy} & 1/m_{xz} \\ 1/m_{yx} & 1/m_{yy} & 1/m_{yz} \\ 1/m_{zx} & 1/m_{zy} & 1/m_{zz} \end{pmatrix} \tag{2.3}$$

to allow for different values along the directions. In general, the mass tensor is an ellipsoid. It is spherical for the electrons at the Γ point with its eigenvalues on the diagonal. This isotropic assumption is a fairly good approximation. For the electron valleys at the L and X points in the Brillouin zone of a cubic crystal, however, the tensor is a spheroid, where the masses are characterized by a longitudinal m_l and two equal transverse masses m_t (see Section 7.2). It is not necessarily diagonal any more. Off-diagonal components are required to describe the electron masses at the L valleys because the principal coordinate system of the corresponding mass tensor does not coincide with the crystal coordinate system. The same can happen for the X valleys if the crystal coordinate system does not coincide with the simulation coordinate system. In this case a rotation of the effective mass tensor is necessary, introducing off-diagonal components in the general case. Several L and X valleys exist, with the spheroidal mass tensor oriented differently for each. Therefore several Schrödinger equations must be solved with identical band edge profiles $V(\mathbf{x})$ but different mass tensors. If the band edges are split due to strain, each valley and thus each Schrödinger equation has a different band profile. If the simulation coordinate system is not identical to the crystal coordinate system, e.g. for growth along the [011] direction, the mass tensor requires an additional

rotation. In wurtzite, a spherical mass tensor is not appropriate at the Γ point. One has to distinguish between the effective mass $m_{\parallel}$ parallel to the hexagonal c axis, and the two masses $m_{\perp}$ perpendicular to it. The inverses of these three masses are the diagonal components of this spheroidal mass tensor.

One can use the same Schrödinger equation for the heavy, light and split-off holes, where the respective heavy, light and split-off valence band edge energy is taken as the potential energy. These energies, including shifts and splittings due to strain, are obtained by diagonalizing for each grid point the bulk 6×6 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian at $\mathbf{k} = 0$, that takes into account the local strain tensor $\epsilon(\mathbf{x})$ (eq. (1.8)) and the deformation potentials of the material at position $\mathbf{x}$. Typically, the single-band model with a spherical mass is not very accurate for the holes, especially if strain is present, because the energy dispersion is not isotropic and instead described by so-called 'warped spheres', see Section 3.1. The masses for the heavy and light holes along specific directions can be obtained from the Luttinger parameters (eq. (3.143) – eq. (3.148)). Nevertheless, this simple model is still very useful for the heavy hole ground state in heterostructures if one is interested in a qualitative picture only, like the variation of the electron–hole interband transition energy as a function of quantum dot radius, electric field or alloy composition, or if the expected simulation time for a $\mathbf{k} \cdot \mathbf{p}$ model is huge. In this case a quick calculation of the single-band Schrödinger equation can be used to test and optimize the design layout before eventually performing the heavy computations.

The discretization of the single-band Schrödinger equation for a heterostructure in real space leads to a real, symmetric eigenvalue problem that is solved iteratively by standard methods (e.g. Arnoldi method (ARPACK) [LSY98]). The dimension of the matrix is equal to the number of grid points. Possible boundary conditions are Dirichlet, Neumann and periodic. If a magnetic field is included (Section 2.3), the eigenvalue problem becomes Hermitian. This is also the case for periodic boundary conditions with nonzero superlattice vectors (Section 2.2).

1D The Schrödinger equation for a semiconductor grown along the z direction and homogeneous along the x and y directions is given by eq. (4.11). Because of the three spatial dimensions one needs three quantum numbers k_x, k_y and n to label the states. The relation $E_n\left(\mathbf{k}_{\parallel}\right)$ (eq. (4.15)) corresponds to a two-dimensional parabola for each n, the so-called electric subband. The one-dimensional single-band Schrödinger equation is given by eq. (4.14). It is also discussed in Section 7.2.

2D The two-dimensional Schrödinger equation is appropriate if the electrons are free to move along the z direction and are confined in the (x, y) plane, e.g. in a quantum wire. The relation $E_n\left(k_{\parallel}\right)$ (eq. (4.37)) now corresponds to a one-dimensional parabola for each n, which is also called electric subband.

Results of single-band calculations are presented in Chapter 7 for silicon, and in Section 10.2 for diamond. The next two sections cover the energy spectra of structures in a uniform magnetic field and the energy spectra of three-dimensional superlattices.

2.2. Periodic boundary conditions – Minibands in artificial quantum dot crystals

Esaki and Tsu proposed in 1969 the concept of a man-made single-crystal with a periodic one-dimensional structural modification, a semiconductor 'superlattice' [ET70]. It is based on a periodic structure of alternating layers of semiconductor materials with wide and narrow band gaps. After the experimental demonstration of one-dimensional superlattices using a sequence of GaAs quantum wells and AlGaAs barriers, it was natural to extend this concept to two-dimensional and three-dimensional superlattices. If the thickness of the barrier layers is small enough, that electrons tunnel through, their wave functions are no longer localized in one quantum well but extend over the whole superlattice. A superlattice is termed an artificial crystal if the electronic wave functions exhibit exactly this tunneling-induced coupling among the superlattice periods. In this section, we investigate the superlattice energy dispersion relation $E_n(\mathbf{K}) = E_n(K_x, K_y, K_z)$ for a regimented ensemble of quantum dots, where n is the miniband index and $\mathbf{K}$ the superlattice vector. Such an arrangement is called artificial quantum dot (QD) crystal. Very remarkably, such three-dimensional structures can nowadays be grown with perfect periodicity by a combined top-down and bottom-up approach using extreme ultraviolet interference lithography, reactive ion etching and templated self-organization [GFD+07]. This templated self-organization of nanostructures is a possible route to achieve exact positioning of quantum dots to create QD arrays, QD molecules and QD crystals. The properties of crystalline solids that can be grown by nature (e.g. NaCl) or in the laboratory (e.g. GaAs) are determined essentially by their interatomic spacing and their symmetry. In contrast, the electronic and optical properties of artificial crystals like 1D quantum well superlattices or QD arrays can be tuned by design and thus offer a lot of freedom in choice of design parameters in order to develop new artificial materials. An artificial crystal is made up from artificial atoms (QDs) (see Fig. 3.7), similarly to the case of a real crystal that is made up from real atoms. If the electronic wave functions of the QDs overlap, extended electron states are formed. Consequently, the individual levels in the QDs are split into bands which are called minibands. These are very similar to the bands in solids. If minibands are formed, not only from a structural point of view but also with respect to the band structure, the 3D ordered quantum dots can be regarded as an artificial crystal where the electrons occupy 'Bloch states'. Future applications of such structures could be solar cells [JG06], high-temperature thermoelectric applications or infrared photodetectors.

Before one analyzes real structures, it is necessary to simulate first simple structures in order to gain confidence in the understanding of the essential features associated to the choice of basic design parameters like QD geometry, QD spacing, QD material parameters, QD asymmetry, etc. Lazarenkova and Balandin [LB01] calculated the miniband dispersion in a three-dimensional QD crystal using an effective mass envelope function approximation. Their crystal is made up of a regular pattern of an array of cubic or tetragonal semiconductor QDs. Their analysis is based on an analytic expression for the confinement potential which is actually an approximation of the real potential. This al-

lows them to separate the three-dimensional Schrödinger equation into three decoupled one-dimensional Schrödinger equations. For a general QD array of arbitrary QD shape, such a simplification is generally not possible. For that reason, we solve numerically the three-dimensional Schrödinger equation. Our approach is so general that it allows for the inclusion of strain, and arbitrary potential variations (e.g. due to QD material composition) automatically. We show that for the special case of cubic QDs with rather high finite barriers, the approach pursued by Lazarenkova and Balandin is reasonably accurate only for the below-the-barrier states. This has been shown by the same authors already in a follow-up publication [LB02], although their previous analytical approach is still used due to its simplicity [JG06].

The coupling of the electronic wave functions of the individual quantum dots leads to a splitting of the energy levels of the single dots and consequently three-dimensional minibands are formed. This electronic band structure can be tuned by e.g. varying QD size, QD spacing, QD materials, QD confinement energies and QD pattern. The properties (e.g. effective mass tensor) derived from this band structure are substantially different from the bulk properties of the involved materials.

2.2.1. Theoretical background

We consider a semiconductor structure that is described by a periodic potential

$$V(x,y,z) = V(x + m_x L_x, y + m_y L_y, z + m_z L_z) = V(\mathbf{x} + \mathbf{L}), \tag{2.4}$$

where L_x, L_y and L_z are the lengths of the superlattice periods along the x, y and z directions, m_x, m_y, m_z are integers and $\mathbf{L}$ is the translation vector. Such a potential occurs for instance in an infinite QD superlattice. The eigenfunctions ψ of a one-electron Hamiltonian with such a periodic potential can be chosen as

$$\psi_{n\mathbf{K}}(\mathbf{x}) = u_{n\mathbf{K}}(\mathbf{x})\, e^{i\mathbf{K}\cdot\mathbf{x}}, \tag{2.5}$$

where it holds

$$u_{n\mathbf{K}}(\mathbf{x} + \mathbf{L}) = u_{n\mathbf{K}}(\mathbf{x}), \tag{2.6}$$

for all superlattice translation vectors $\mathbf{L}$ (Bloch's theorem or Bloch–Floquet theorem). $\psi_{n\mathbf{K}}$ is the Bloch function of band index n and consists of the product of a plane wave term $e^{i\mathbf{K}\cdot\mathbf{x}}$ and the periodic function $u_{n\mathbf{K}}$ which has the same periodicity as the potential V. Eq. (2.5) and eq. (2.6) imply

$$\psi_{n\mathbf{K}}(\mathbf{x} + \mathbf{L}) = \psi_{n\mathbf{K}}(\mathbf{x})\, e^{i\mathbf{K}\cdot\mathbf{L}}. \tag{2.7}$$

The periodicity requires that the physical characteristics of the system do not change if the electron is shifted exactly by integer number of periods. The effect of a lattice translation is to change the wave function only by a phase factor $e^{i\mathbf{K}\cdot\mathbf{L}}$. $\mathbf{K}$ denotes the superlattice wave vector of the electron moving in this potential and is a real number. By choosing macroscopic periodic (Born–von Karman or cyclic) boundary conditions which implies that the wave functions must be periodic with respect to the crystal boundaries

$$\psi(x + M_x L_x, y + M_y L_y, z + M_z L_z) = \psi(x,y,z), \tag{2.8}$$

and applying Bloch's theorem (eq. (2.7)) to this equation, the number of allowed values for $\mathbf{K}$ will be restricted by the requirement

$$e^{i(M_x K_x L_x + M_y K_y L_y + M_z K_z L_z)} = 1. \tag{2.9}$$

Here, the total number of superlattice unit cells in the crystal is denoted by $M = M_x M_y M_z$ where M_i are all integers. This leads to the allowed values of $\mathbf{K}$

$$K_i = \frac{m_i}{M_i} \frac{2\pi}{L_i} \qquad (i = x, y, z)\,, \tag{2.10}$$

where the spacing between allowed value of K_i is given by $\frac{2\pi}{M_i L_i}$. For large values of M_i the discreteness of the allowed values of K_i becomes negligible. The charge carriers are equally likely to be found in any superlattice period. The energy spectrum is a periodic function of $\mathbf{K}$ with period $2\pi/L_i$, thus it is sufficient to restrict oneself without loss of generality to the first superlattice Brillouin zone $[-\pi/L_i, \pi/L_i]$. For $\mathbf{K} = 0$ or $K_i = \pm\frac{\pi}{L_i}$, the Bloch wave is stationary. Usually one only plots the interval $[0, \pi/L_i]$ because the parity property

$$E_n(\mathbf{K}) = E_n(-\mathbf{K})\,, \tag{2.11}$$

is fulfilled by any superlattice band. Technically, for the numerical simulations with next**nano**[3] one specifies the length of the superlattice periods L_i and the number of superlattice vectors. These two quantities then determine the actual crystal size, i.e. the number of identical primitive superlattice unit cells. For instance, if one specifies 20 superlattice vectors for the interval $[-\pi/L_x, \pi/L_x]$ (ignoring counting the Gamma point $\mathbf{K} = 0$), the total crystal contains 20 superlattice unit cells along the direction x, and has a total length of $20L_x$. It is important to remember that the number of superlattice vectors is not only a property that reflects the 'grid resolution' of the miniband dispersion in $\mathbf{K}$ space, but that it actually corresponds to the number of repeated unit cells. In real samples where the number of unit cells is not infinite but has a finite value, the discreteness of the miniband dispersion might be relevant for evaluating the density of states. $\mathbf{K}$ is actually a quantum number of the system. It characterizes the wave functions corresponding to different states of the same miniband.

2.2.2. Miniband dispersion in a bulk 1D sample

We consider a simple, trivial example where we take a bulk GaAs sample of length $L_x = 12\,\text{nm}$ which corresponds to our superlattice unit cell. We apply periodic boundary conditions and solve the single-band effective-mass Schrödinger equation for the electrons for each of the 20 superlattice vectors in the interval between $[-\pi/L_x, \pi/L_x]$. Our choice of discreteness of the allowed values of K_x determines the sample size ($20L_x = 240\,\text{nm}$). We plot the energy dispersion for the lowest three minibands. They arise from the first, second and third eigenstate of the 12 nm wide superlattice unit cell. The energies of the second and third eigenstate are degenerate at $K_x = 0$, whereas at $K_x = -\pi/L_x$ and $K_x = \pi/L_x$, the first and second eigenstate are degenerate. It follows, that there are no band gaps within this miniband spectrum. As we simulated only plain GaAs, the

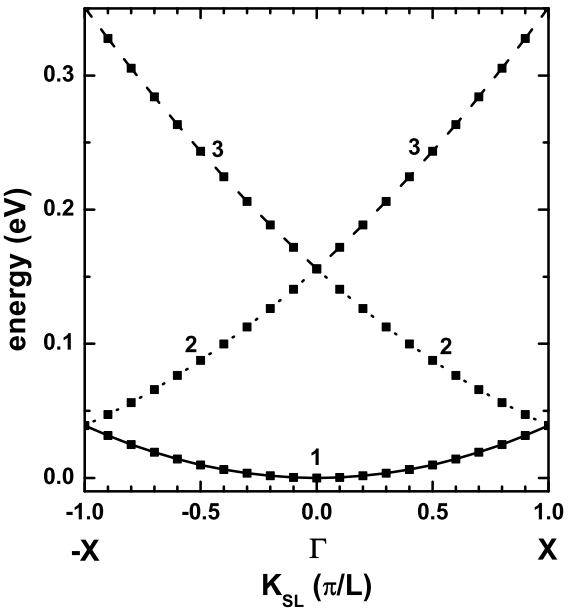


Figure 2.1.: Electron energy dispersion of the three lowest minibands of a simple bulk GaAs sample of periodic length $L_x = 12\,\mathrm{nm}$. The conduction band edge is set to 0 eV. The lines are a guide to the eye.

dispersion of these three minibands must follow the parabolic energy dispersion of bulk GaAs, i.e. $E(k) = \frac{\hbar^2 k^2}{2m}$, where $m = 0.067 m_0$ is the conduction band effective mass of GaAs. For a very large number of superlattice vectors, i.e. for a very large bulk sample, the minibands form a continuous spectrum, starting from the conduction band edge.

Figure 2.2 shows the energy levels of the three lowest eigenstates E_i and their corresponding probability density (ψ_i^2) (shifted by their eigenenergies E_i) for the superlattice vector $K_x = 0$ (stationary Bloch states). The square of the ground state wave function ψ_1^2 is constant with its energy $E_1 = 0\,\mathrm{eV}$ equal to the conduction band edge energy. The energies of the second (E_2) and third (E_3) eigenstate are degenerate at $K_x = 0$. Their wave functions have a cosine and sine shape, respectively.

2.2.3. Miniband dispersion in Ge–Si quantum dot crystals

As a second example, we calculate the miniband dispersion of a cubic Ge–Si QD crystal by numerically solving the three-dimensional Schrödinger equation on a finite-differences grid. The QD dimensions are $L_x = L_y = L_z = 6.5\,\mathrm{nm}$. The barrier thickness is chosen to be $H_x = H_y = H_z = 1.5\,\mathrm{nm}$. The imposed QD periodicity is thus given by $d_i = L_i + H_i = 8.0\,\mathrm{nm}$ ($i = x, y, z$). We solve the single-band Schrödinger equation

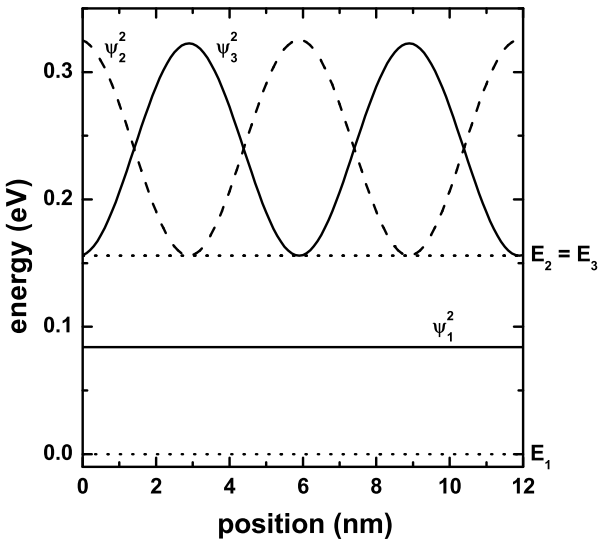


Figure 2.2.: Stationary Bloch states of a 12 nm bulk GaAs sample with periodic boundary conditions. Shown are the three lowest energy levels E_1, E_2 and E_3, as well as their associated probability densities. The conduction band edge is set to 0 eV.

for the holes assuming an isotropic effective mass tensor. The mass for the QD is $m_{\mathrm{w}} = m_{\mathrm{Ge}} = 0.28m_0$ and for the surrounding barrier material $m_{\mathrm{b}} = m_{\mathrm{Si}} = 0.49m_0$. The valence band offset is taken to be $E_{\mathrm{VBO}} = 0.45\,\mathrm{eV}$. The choice of material parameters is based on Ref. [LB01] and corresponds roughly to the heavy hole states in a Ge–Si QD crystal. In this simple example, we neglect the anisotropic nature of the hole states, strain and the interaction between heavy, light and split-off hole states. Later, the valence band anisotropy of a spherical Si QD is discussed in Section 3.4, as well as the influence of spin-orbit coupling. We solve the 3D Schrödinger equation (eq. (2.2)) in real space where we impose periodic boundary conditions along the x, y and z directions, i.e. we numerically discretize only a single QD on a rectangular grid having the grid spacing 0.50 nm (i.e. 16 grid points in each direction). Thus the size of the Schrödinger matrix to be solved is $16 \times 16 \times 16 = 4096$. This Schrödinger equation has to be solved for different values of the superlattice vector $\mathbf{K} = (K_x, K_y, K_z)$. For $\mathbf{K} = 0$ or $K_i = \pm\frac{\pi}{L_i}$, where the Bloch wave is stationary, the matrix eigenvalue problem is symmetric and real. In contrast, it is Hermitian for all other nonzero superlattice vectors $\mathbf{K}$. With a grid resolution of 0.50 nm for QDs of such size, the expected error compared to an extremely dense grid has been found to be of the order $1\% - 4\%$ [LB02]. In this example, we restrict ourselves to calculating the energy dispersion along special directions in $\mathbf{K}$ space only, i.e. from $\Gamma_{\mathrm{SL}} = (0,0,0)$ to $\mathrm{X}_{\mathrm{SL}} = \mathbf{K}_{100} = (1,0,0)$, from $\Gamma_{\mathrm{SL}} = (0,0,0)$ to $\mathrm{K}_{\mathrm{SL}} = \mathbf{K}_{110} = (1,1,0)$

and from $\Gamma_{\mathrm{SL}} = (0,0,0)$ to $\mathrm{L}_{\mathrm{SL}} = \mathbf{K}_{111} = (1,1,1)$, where Γ_{SL}, X_{SL}, K_{SL} and L_{SL} are the names for the special points in $\mathbf{K}$ space. The directions labeled by [100], [110], [111] refer to the simulation coordinate system and not to the crystal coordinate system.

The following three figures show the calculated heavy hole energy dispersion of an artificial cubic Ge–Si quantum dot crystal along the [100] (Fig. 2.3), [110] (Fig. 2.4(a)) and [111] (Fig. 2.4(b)) quasicrystallographic directions with the parameters $L_x = L_y = L_z = 6.5\,\mathrm{nm}$ and $H_x = H_y = H_z = 1.5\,\mathrm{nm}$. The minibands are labeled by three quantum numbers n_x, n_y, n_z which indicate the symmetry of the corresponding wave functions. This labeling is strictly speaking only correct in the case of separable wave functions along the x, y and z directions. The superscript indicates the number of degeneracy. The right part of Fig. 2.3 shows schematically the valence band edge profile of the QD and the position of the energy levels for $\mathbf{K} = 0$. The figures show the energy spectrum of the lowest 50-60 eigenstates (neglecting spin degeneracy) with respect to the heavy hole valence band edge. The energy scale has been chosen such that the maximum valence band edge energy of the Ge QD is at $E_{\mathrm{v,Ge}} = 0.45\,\mathrm{eV}$, and the maximum valence band edge energy of the Si barrier is at $E_{\mathrm{v,Si}} = 0\,\mathrm{eV}$ which is indicated by the dashed line. The number of superlattice vectors along the directions in $\mathbf{K}$ space determines the resolution of the energy dispersion plots. We used 11 superlattice vectors in these plots, i.e. the Schrödinger equation had to be solved 11 times. Our results agree very well with the analytical calculations of Ref. [LB01] in the energy region where the confinement inside the QD is strong. For the higher lying states inside the QD and above the barrier, our results differ because we use the correct 3D QD confinement potential, whereas in Ref. [LB01] the potential landscape has been approximated with an analytical ansatz that allows for the separation of the x, y and z variables. We want to point out that this ansatz is only justified for states confined deep inside the QD. For states above the barrier, this ansatz will lead to significant (i.e. three times the valence band offset energy) artificial potential barriers at the positions where the three independent potential barriers overlap. We thus believe that one of the conclusions of Ref. [LB01], namely that the 3D regimentation of quantum dots in QD crystals leads to the appearance of 'resonant' quasidiscrete energy levels above the potential barrier for large interdot distances is a pure artefact of this separation ansatz. At $\mathbf{K} = 0$, the degeneracy is higher than at nonzero $\mathbf{K}$ vectors where the symmetry in the superlattice Brillouin zone is lower. Note that the eigenstates along the [111] direction show a higher degree of degeneracy throughout the superlattice Brillouin zone as compared to [100] and [110]. Both, the QD itself and the QD superlattice have the same cubic symmetry in this example. Thus the degeneracy of the 123 (including permutations) energy band is sixfold throughout the Brillouin zone along the [111] directions (Fig. 2.4(b)). Since the tunneling-induced coupling increases with miniband index (i.e. energy), the width of the bands with respect to energy also increases. Consequently, the higher bands must overlap and the narrowest band is the ground state band. An important feature are the band gaps in the dispersion curve which define the energy intervals in which propagating states do not exist. For a detailed discussion of these dispersion curves and their degree of degeneracy we refer to Ref. [LB01].

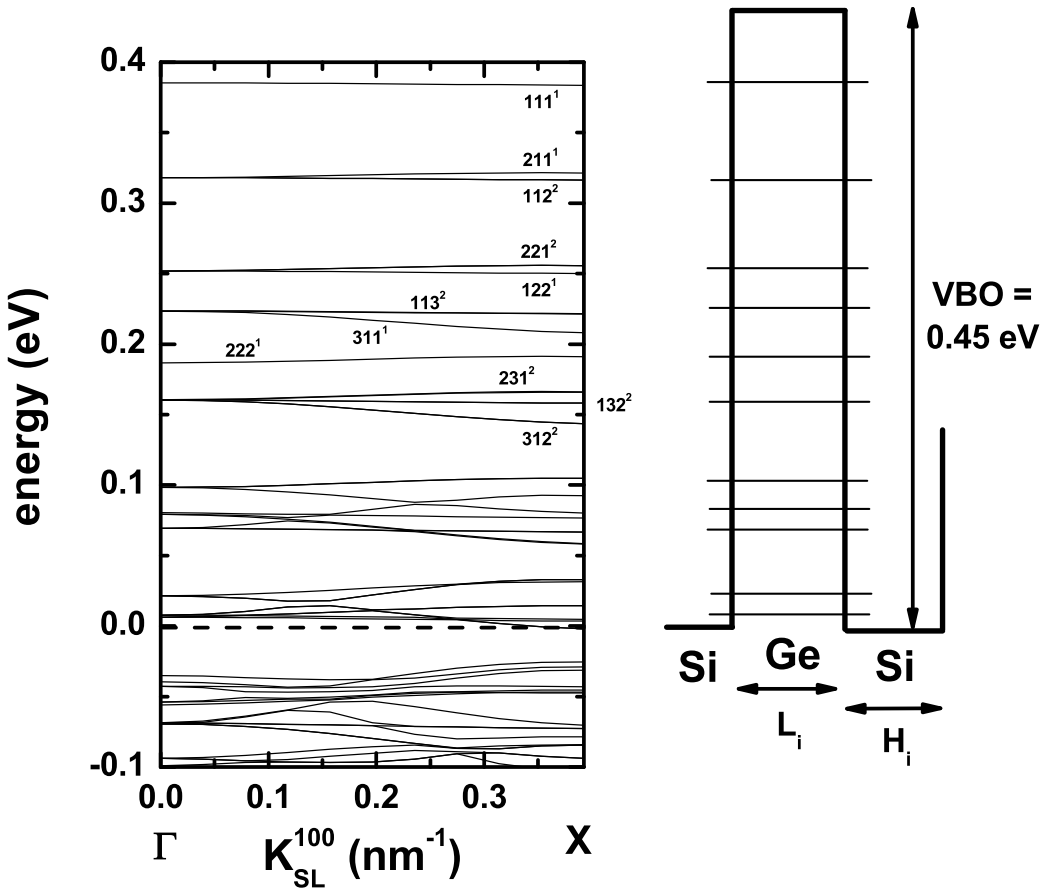
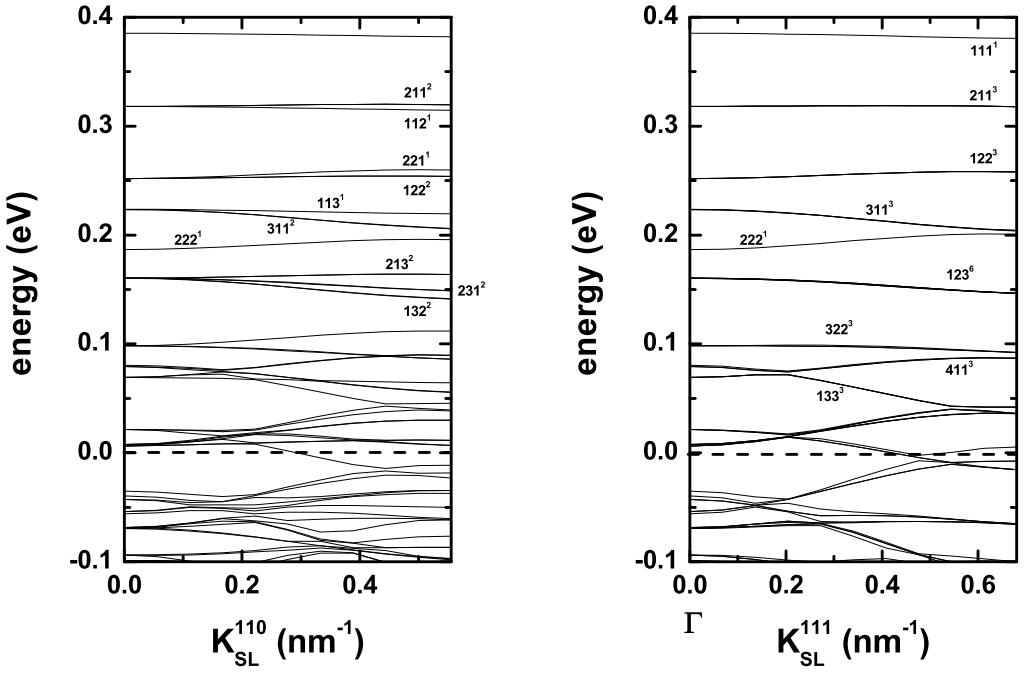


Figure 2.3.: Heavy hole energy dispersion of an artificial cubic Ge–Si quantum dot crystal along the [100] quasicrystallographic direction with the parameters $L_x = L_y = L_z = 6.5\,\mathrm{nm}$, $H_x = H_y = H_z = 1.5\,\mathrm{nm}$. The valence band edge of the Ge dot is at 0.45 eV, the valence band edge of the Si barrier is at 0 eV which is indicated by the dashed line. The minibands are labeled by the three quantum numbers n_x, n_y, n_z. The right part shows schematically the valence band edge profile of the QD and the position of the energy levels for $\mathbf{K} = 0$.



(a) [110] direction

(b) [111] direction

Figure 2.4.: Heavy hole energy dispersion of the artificial QD crystal of Fig. 2.3 along the (a) [110] and (b) [111] quasicrystallographic directions.

Now we change the QD shape, as well as the shape of the repeating pattern, from a cubical to a tetragonal shape, analogous to Ref. [LB01]. The parameters are $L_x = L_y = 5\,\text{nm}$, $L_z = 2.5\,\text{nm}$ and $H_x = H_y = 2.5\,\text{nm}$, $H_z = 1.25\,\text{nm}$. The superlattice periods are $d_x = d_y = 7.5\,\text{nm}$ and $d_z = 3.75\,\text{nm}$. The grid spacing is 0.25 nm in all three directions. The resulting Schrödinger matrix has a dimension of $30 \times 30 \times 15 = 13500$. All other assumptions are the same as for the above cubic QD superlattice example. The following three figures show the calculated dispersion relation along the [100] (Fig. 2.5), [110] (Fig. 2.6(a)) and [111] (Fig. 2.6(b)) quasicrystallographic directions. Our results are in excellent agreement to the numerical results of [LB02].

The above calculations have been performed for an ideal model QD system in order to benchmark our results with previous publications. The algorithm that we implemented into the next**nano**^{3} software is so general that arbitrary QD geometries, strain, doping and different crystallographic orientations, for instance, can be treated as well.

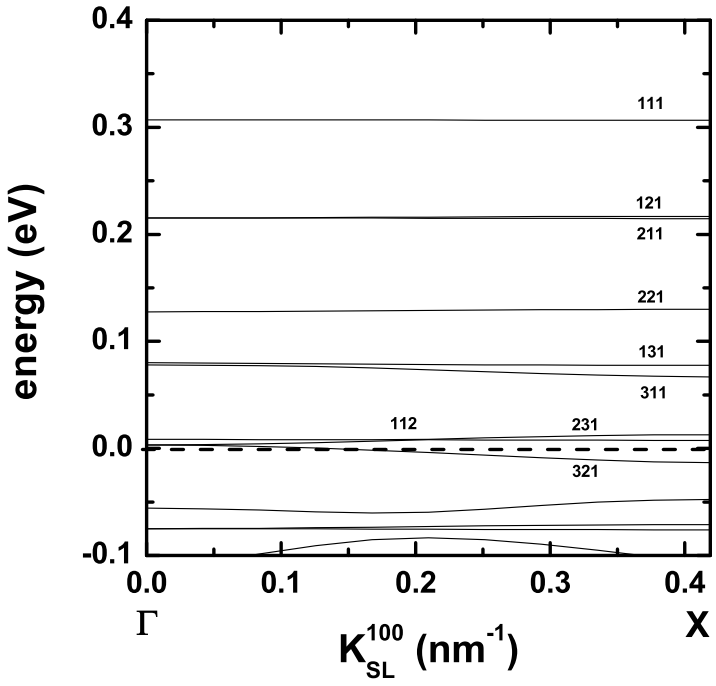
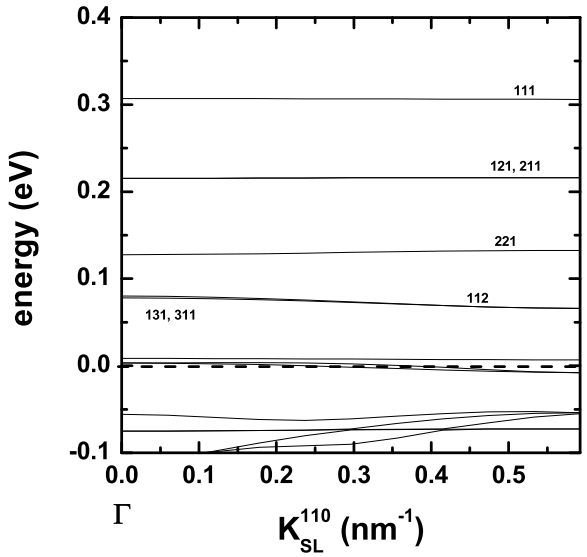
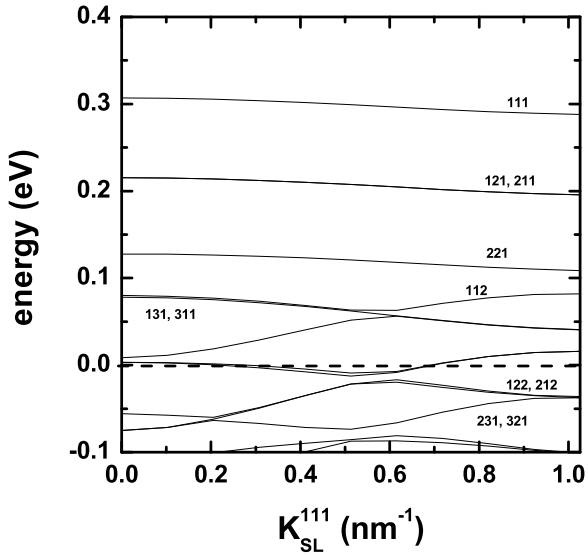


Figure 2.5.: Heavy hole energy dispersion of an artificial tetragonal Ge–Si quantum dot crystal along the [100] quasicrystallographic direction with the parameters $L_x = L_y = 5\,\text{nm}$, $L_z = 2.5\,\text{nm}$, $H_x = H_y = 2.5\,\text{nm}$, $H_z = 1.25\,\text{nm}$. The valence band edge of the dot is at 0.45 eV, the valence band edge of the barrier is at 0 eV which is indicated by the dashed line. The minibands are labeled by the three quantum numbers n_x, n_y, n_z.



(a) [110] direction



(b) [111] direction

Figure 2.6.: Heavy hole energy dispersion of an artificial tetragonal Ge–Si quantum dot crystal along the (a) [110] and (b) [111] quasicrystallographic directions. The QD crystal is the same as in Fig. 2.5.

2.3. Magnetic field

An important but sometimes ignored aspect of the numerical solution of the Schrödinger equation in a magnetic field is gauge invariance. Since the vector potential increases linearly as a function of position, its values tend to be one or two orders of magnitude larger at the boundaries of the simulation region than in the center. Therefore, the numerical solutions depend sensitively on the chosen gauge for nonzero magnetic field, even if the zero field wave functions are small at the boundaries. This problem was successfully addressed in Ref. [GU98] for single-band, nonrelativistic Hamiltonians. Andlauer et al. [AMV08, TZA+06] recently generalized this scheme for arbitrary envelope function Hamiltonians in a way that ensures manifestly gauge invariant results for any magnetic field.

2.3.1. Fock–Darwin spectrum

Quantum dots (QDs) that are subject to a magnetic field are an interesting research subject since many years. A popular approach to study the energy spectrum of such systems theoretically is the assumption of a two-dimensional parabolic confinement potential that is subject to a perpendicularly oriented magnetic field which introduces a further parabolic confinement, thus making it possible to solve this equation analytically [KAT01]. Obviously, such an approach only makes sense for self-assembled quantum dots that have a strong confinement in the growth direction z and a weak (roughly parabolic) confinement in the plane perpendicular to it. However, this model might not be applicable to all QD systems, such as spherical QDs, where the above mentioned two-dimensional model neglects completely the quantum confinement along the third direction. Consequently, it would be desirable to have a more predictive model that takes into account the realistic three-dimensional potential profile of quantum dots of arbitrary shape, including strain, deformation potentials and piezoelectric effects, and then apply the magnetic field and calculate the energy levels. Governale et al. [GU98] showed how to discretize the single-particle Schrödinger equation on a two-dimensional homogeneous grid in a gauge-invariant way assuming a constant effective mass tensor. We extended their method to spatially varying effective masses and to an inhomogeneous grid [Hac02] and are now able so solve the Schrödinger equation in two and three dimensions including a uniform magnetic field. The Hamilton operator of eq. (2.1) has to be modified to include the magnetic field and is now given by

$$\mathbf{H} = -\frac{\hbar^2}{2}\left((\nabla + ig\mathbf{A}) \cdot \mathbf{M}\left(\mathbf{x}\right)(\nabla + ig\mathbf{A})\right), \tag{2.12}$$

where $g = e/\hbar$ and $\mathbf{A}(\mathbf{x})$ is the vector potential which is discretized in the symmetric gauge

$$\mathbf{A}(\mathbf{x}) = -\frac{1}{2}\mathbf{x} \times \mathbf{B}. \tag{2.13}$$

As a benchmark for our method, we chose to reproduce the well-known Fock–Darwin spectrum [KAT01] which is an analytical result. We note that our result is completely

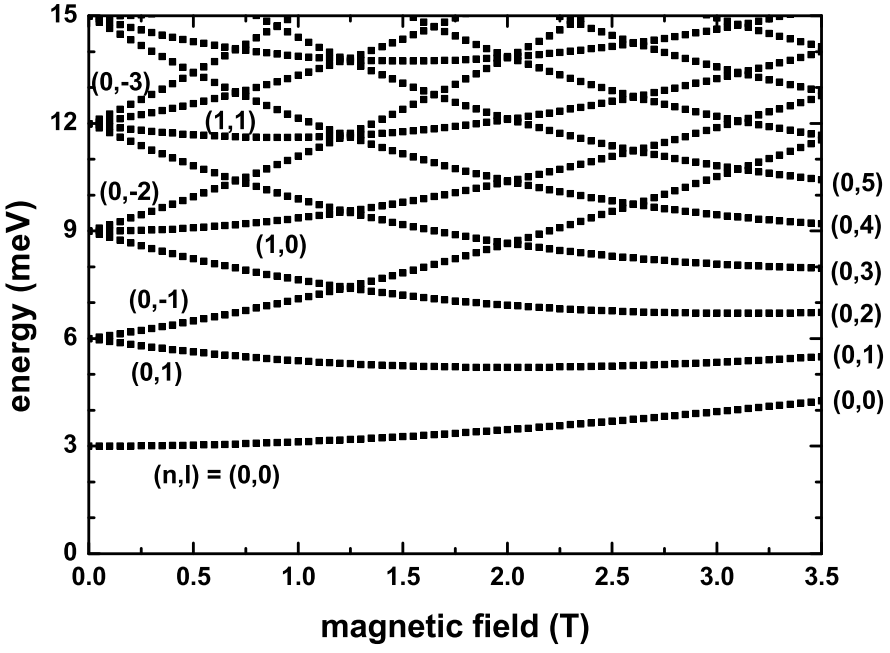


Figure 2.7.: Calculated single-particle energy levels of a two-dimensional parabolic confinement potential ($\hbar\omega_0 = 3\,\mathrm{meV}$) as a function of magnetic field (Fock–Darwin spectrum). The states are labeled by (n, l) which refers to the radial quantum number n and the angular momentum quantum number l.

numerical, i.e. we solve the single-particle 2D Schrödinger equation in the (x, y) plane for a two-dimensional parabolic confinement potential where the uniform magnetic field is applied along the z direction. The parabolic confinement had been chosen so that the energy separation between the states is $\hbar\omega_0 = 3\,\mathrm{meV}$ (without magnetic field) where ω_0 is the oscillator frequency of the parabolic confinement. The effective mass tensor has been assumed to be isotropic and constant ($m = 0.067 m_0$, electron effective mass of GaAs). Without magnetic field, and neglecting the twofold spin degeneracy, the ground state is not degenerate, the second level is twofold degenerate, the third level threefold, and so on, as can be seen from the analytical result of the energy spectrum

$$E_{n,l} = (2n + |l| - 1)\,\hbar\omega_0, \tag{2.14}$$

where $n = 1, 2, 3, ...$ is the radial quantum number and $l = 0, \pm 1, \pm 2, ...$ the angular momentum quantum number. If the magnetic field is present, the eigenenergies are

$$E_{n,l} = (2n + |l| - 1)\,\hbar \left(\omega_0^2 + \frac{1}{4}\omega_\mathrm{c}^2\right)^{1/2} - \frac{1}{2} l \hbar \omega_\mathrm{c}, \tag{2.15}$$

where $\omega_c = eB/m$ is the cyclotron frequency. Thus the degeneracy of the 2D harmonic oscillator is lifted as can be seen in the calculated energy spectrum (Fig. 2.7), where the energy levels are plotted as a function of the magnetic field magnitude.

A more detailed discussion of this energy spectrum can be found in Ref. [KAT01]. We conclude that our numerical calculations are in perfect agreement to the analytical results (not shown), and therefore, our method can be straightforwardly applied to realistic, three-dimensional confinement potentials to obtain the energy spectrum of the transitions of quantum dots subject to a magnetic field.

2.3.2. Results: Qubit manipulation

We now apply this method to coupled quantum wires in a longitudinal magnetic field [BZA+07]. Two GaAs quantum wells, each with a width of 14.5 nm, are vertically stacked along the y direction and are separated by a thin $Al_{0.32}Ga_{0.68}As$ tunnel barrier of width 1 nm. Electrons are provided by two δ-doped layers on each side of the double quantum well structure ($\delta_{top} = 2.1 \cdot 10^{12}$ cm^{-2}, $\delta_{bottom} = 2.9 \cdot 10^{11}$ cm^{-2}). The upper well is 60 nm below the surface where we pin the Fermi energy by surface states (see Appendix A) at the middle of the band gap. A top gate voltage of 0.04 V is applied to this Schottky contact leading to nearly symmetric quantum wells. The top gate voltage can be used to tune the electron density, i.e. to shift the ground state from the top to the bottom well (the first excited state is then located in the opposite well) or to vary the degree of wave function mixing (bonding–antibonding) among the two wells. The actual quantum wires are obtained by wet-etching nanogrooves into the semiconductor surface which locally deplete the two-dimensional electron gas of the GaAs–AlGaAs double quantum well heterostructure and produce two electrostatically defined quantum wires on top of each other (Fig. 2.8(c)). The experimental details are described in Ref. [FAK+06]. Such wires are short, ballistic 1D electron systems. (Ballistic transport is covered in Chapter 4.) They recently attracted attention because controllable coupling can be achieved, making such devices interesting for solid-state quantum information processing. Superposition states are formed by tunnel coupling between the quantum wires. They can be altered by an applied magnetic field along the wire direction leading to tunable wave function mixing, in addition to the mixing produced by the top gate. To implement such quantum logic devices in quantum circuits, one would like to understand and have detailed control over all involved electron levels which calls for a realistic modeling such as provided by next**nano**. We have solved the 2D Schrödinger–Poisson equation of the double quantum well heterostructure for electrons self-consistently, taking into account the spatially varying material parameters. We have modeled the confinement due to the nanogrooves by a parabolic potential along the x direction and set its characteristic energy such that it matches the observed low-lying level spacing of 10 meV.

For the resulting 2D confinement potential (Fig. 2.8(a)), we have calculated the electronic eigenstates as a function of a magnetic field oriented along the wire (z) direction, i.e. perpendicular to the 2D confinement potential landscape. For the conduction band offset between GaAs and $Al_{0.32}Ga_{0.68}As$ we have used a value of 0.30 eV.

Our numerical calculations of the energy spectrum (Fig. 2.9) reproduce the experi-

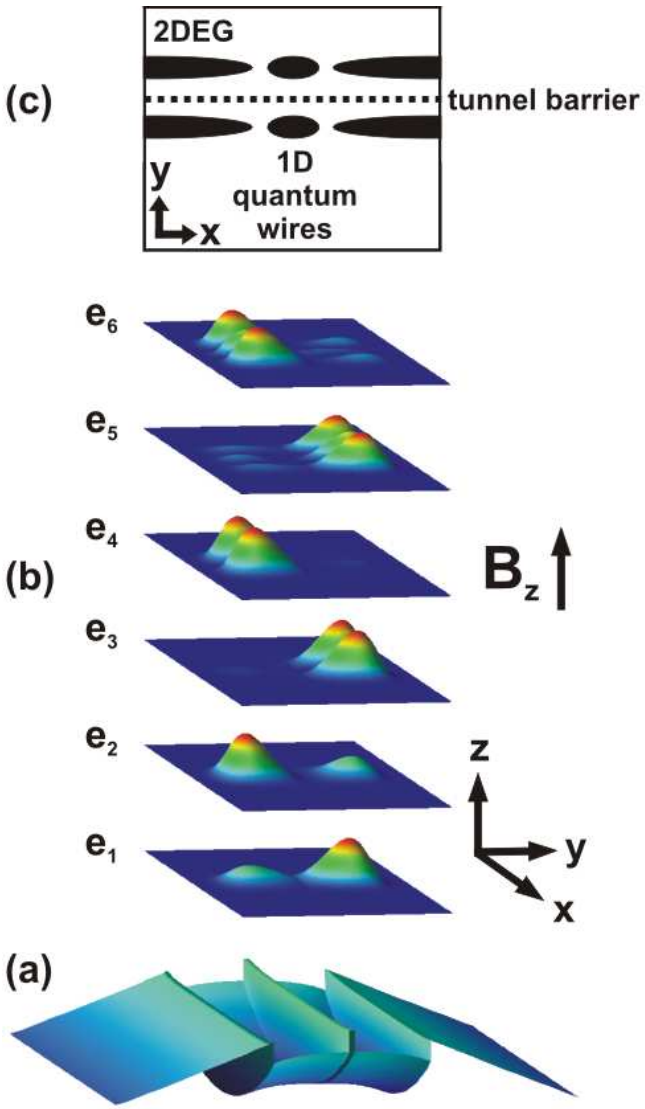


Figure 2.8.: (a) Two-dimensional conduction band profile $E_c(x, y)$ of the coupled quantum wires in the (x, y) plane with z = constant, (b) square of the wave functions $|\Psi_i(x, y)|^2$ for the six lowest electron eigenstates at a magnetic field of 4.5 T oriented along the wire (z) direction, (c) schematic cross section of the electron systems (dark), the 1D quantum wires are formed by electrostatically depleting the two-dimensional electron gas (2DEG).

mental transconductance maxima (compare with Fig. 8(b) of Ref. [FAK$^+$06]) very well. These maxima directly image the subband edges of one-dimensional ballistic electron transport in electron waveguides. One can clearly see level anticrossings to occur at certain magnetic fields. For those energies, the tunneling coupling is significantly reduced or

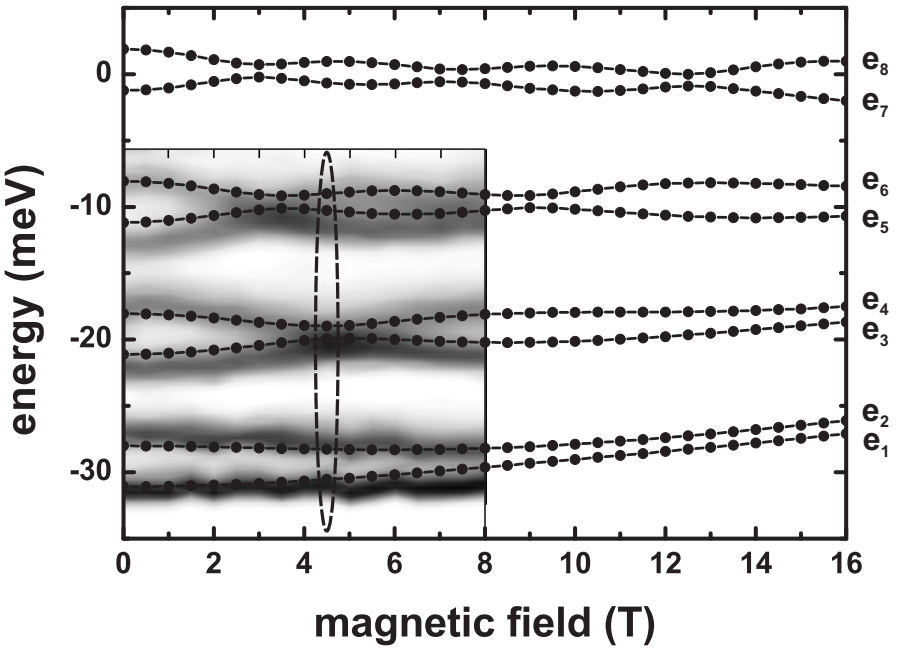


Figure 2.9.: Calculated energy spectrum of the eight lowest electron eigenstates of the coupled quantum wires as a function of magnetic field strength. (The lines are a guide to the eye.) The dashed ellipse highlights the energies of the six eigenstates at 4.5 T that are plotted in Fig. 2.8(b). For comparison, the rectangle shows the gray-scale plot of the transconductance maxima versus top gate voltage and magnetic field from 0 T to 8 T (experimental data after Fig. 8(b) of Ref. [FAK$^+$06]). The top gate voltage was scaled so that the subband energies approximately align with the energy scale.

vanishes. This is the case, for example, for the third and forth eigenstate in Fig. 2.9 for a magnetic field strength of 4.5 T. Consequently, the probability densities of these states (labeled e_3 and e_4 in Fig. 2.8(b)) are not smeared out any longer over the two wires, in contrast to the other states which still show wave function tunneling. Such a structure can be considered a qubit with logic states that are represented by the presence of the electron in the upper or lower quantum wire, respectively, and a coupling window that allows for electron transfer between these wires. By varying the magnetic field, one is able to switch between the different logic states. Note that the energy spectrum shown in Fig. 2.9 does not include the (very small) spin splitting. Our results agree nicely qualitatively with analytic studies [MSF07] that ignored charge redistribution due to doping and applied gate voltages. An important difference is the fact that we find anti-crossing behavior, whereas the analytical results always yield a crossing behavior which is equivalent of assuming that the tunnel splitting vanishes completely for some fields. The challenge in this concept is to grow symmetric quantum wire structures in order to reduce the tunnel splitting as much as possible.

3. Multi-band $\mathrm{k} \cdot \mathrm{p}$ envelope function approximation

3.1. The multi-band $\mathrm{k} \cdot \mathrm{p}$ Schrödinger equation

In this section we focus on the $\mathbf{k} \cdot \mathbf{p}$ method but do not consider strain. The effect of strain on the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian has been described in detail in Ref. [Hac02]. We describe the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian for a bulk semiconductor but use the notation suited for heterostructures. In subsequent sections we present results on heterostructures, like spurious solutions and intersubband transitions in multi-quantum well structures. We also model heterostructures showing the untypical type-II and type-III band alignments and compare our results to tight-binding calculations. In Section 10.2, we analyze self-consistent $\mathbf{k} \cdot \mathbf{p}$ calculations of a two-dimensional hole gas in diamond for different substrate orientations. Results of $\mathbf{k} \cdot \mathbf{p}$ calculations on bulk semiconductors are presented in Section 3.6 for unstrained and strained InAs, and in Section 10.1 for diamond. Also these results are compared to tight-binding calculations. The main emphasis of this section, however, is to pull together all the various definitions of the $\mathbf{k} \cdot \mathbf{p}$ parameters available in the literature, and provide equations that relate them to each other, for both zinc blende and wurtzite crystals. We believe that such a summary on a few pages is very useful, not only for beginners but also for experienced researchers that quickly want to look up conversion formulas. However, it turns out that we are not the first ones to write up such a collection. Meanwhile, the recently published book by Lew Yan Voon and Willatzen [LYVW09] offers an even more detailed description and comprehensive overview on all the different $\mathbf{k} \cdot \mathbf{p}$ Hamiltonians used in the literature for both bulk and nanostructured semiconductors.

Compared to the single-band model, a more accurate description of the band structure can be obtained by using the multi-band $\mathbf{k} \cdot \mathbf{p}$ method that has been used since the 1950s. There are a number of different $\mathbf{k} \cdot \mathbf{p}$ models discussed in the literature, e.g. Refs. [CP66, RAF04, PZ96, BRRB$^+$11, CC92, Bah90]. They differ mainly in the number of bands considered and their treatment of strain and spin-orbit interaction. The model that considers up to 14 bands predicts almost perfectly the bulk band structure but the computational effort turns out to be nearly as large as for the empirical tight-binding approach. In next**nano** we make use of the 8-band model (e.g. [Bah90]). It is a compromise between the accuracy and the computational cost, as well as the number of required (and known) material parameters. This model includes the lowest conduction band and the three highest valence bands. All other remote bands are treated as perturbations. Spin-orbit interaction and strain are taken into account as small perturbations.

There are basically two ways in obtaining the bulk $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian matrix, the

first one is based on a perturbative approach pioneered by Dresselhaus, Kip and Kittel [DKK55], the second one is based on symmetry analysis (method of invariants) introduced by Luttinger [Lut56]. A few years ago, Foreman derived six- [For93] and eight-band Hamiltonians [For97] from Burt's exact envelope function theory [Bur92, Bur99] for heterostructures. He showed that his nonsymmetrized Hamiltonian for a homogeneous infinite sample is consistent to the bulk $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, and that deriving the heterostructure Hamiltonian from the bulk one using a symmetrization procedure is incorrect (symmetrized Hamiltonian). These works solved the problem of operator ordering. The reason lies in the noncommutativity of the differential operator and the (position dependent) material parameters. We note that the potential energy term of Burt's exact envelope function equation contains an extra nonlocal term ($V_{nm}(\mathbf{x},\mathbf{x}')$) that has been neglected. In fact, using some approximations, it can be shown that the nonlocal part does not contribute for slowly varying envelope functions. At distances far away from a heterointerface the potential tends to a constant, the local periodic potential, and the nonlocal contribution is small. Close to heterointerfaces, Burt's theory leads to two correction terms to the potential function. Another view is that perturbative effects of material inhomogeneities lead to so-called interface Hamiltonians. A detailed derivation and discussion of the Burt–Foreman theory is given in Ref. [LYVW09].

The key feature of the $\mathbf{k}\cdot\mathbf{p}$ method is the envelope function ansatz based on Bloch's theorem, according to which the electron wave function in a crystal with translational symmetries can be separated into an oscillating Bloch part which is periodic over atomic distance and a smooth envelope function which varies on a mesoscopic scale. Using Löwdin perturbation theory the rapidly oscillating Bloch functions can be eliminated from the electron Hamiltonian. Thus the resulting electron Hamiltonian only contains the envelope functions. A detailed summary of the $\mathbf{k}\cdot\mathbf{p}$ method has been presented in Refs. [Zib07, And09].

Compared to the single-band Schrödinger equation (eq. (2.2)), we now consider an additional term $\mathbf{H}_{\mathrm{so}}$ that approximately takes into account the relativistic effect of spin

$$\mathbf{H}_{\mathrm{so}} = \frac{\hbar^2}{4m_0^2c^2}\left(\nabla V\times\mathbf{p}\right)\cdot\sigma, \tag{3.1}$$

where V is the potential energy term, $\mathbf{p}$ is the momentum operator and σ is the vector of the Pauli matrices $\sigma=(\sigma_1,\sigma_2,\sigma_3)^{\mathrm{T}}$. The one-electron Schrödinger equation now reads

$$\left(\mathbf{H}_0+\mathbf{H}_{\mathrm{so}}\right)\Psi_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\,\Psi_{n\mathbf{k}}(\mathbf{x}) \tag{3.2}$$

$$\left(\frac{\mathbf{p}^2}{2m_0}+V(\mathbf{x})+\frac{\hbar^2}{4m_0^2c^2}\left(\sigma\times\nabla V\right)\cdot\mathbf{p}\right)\Psi_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\,\Psi_{n\mathbf{k}}(\mathbf{x}), \tag{3.3}$$

where $\Psi_{n\mathbf{k}}$ is the Bloch function

$$\Psi_{n\mathbf{k}}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}}u_{n\mathbf{k}}(\mathbf{x}), \tag{3.4}$$

composed of the product of a plane wave $e^{i\mathbf{k}\cdot\mathbf{x}}$ and the periodic Bloch factors $u_{n\mathbf{k}}(\mathbf{x})$. n is the band index and $\mathbf{k}$ is a wave vector in the first Brillouin zone which corresponds

to the periodicity of the potential energy $V(\mathbf{x})$. If we insert the Bloch function $\Psi_{n\mathbf{k}}(\mathbf{x})$ into eq. (3.3), we obtain after canceling the plane wave $e^{i\mathbf{k}\cdot\mathbf{x}}$

$$\left(\mathbf{H}_0 + \mathbf{H}_{\mathbf{k}} + \mathbf{H}_{\mathbf{k}\cdot\mathbf{p}} + \mathbf{H}_{\mathrm{so}}\right) u_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\, u_{n\mathbf{k}}(\mathbf{x}) \tag{3.5}$$

$$\left(\frac{\mathbf{p}^2}{2m_0} + V(\mathbf{x}) + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar}{m_0}\mathbf{k}\cdot\mathbf{p} + \frac{\hbar^2}{4m_0^2c^2}\left(\nabla V\times\mathbf{p}\right)\cdot\sigma\right) u_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\, u_{n\mathbf{k}}(\mathbf{x}), \tag{3.6}$$

which is now written for the periodic Bloch spinor $u_{n\mathbf{k}}(\mathbf{x})$ only. Within our approximation we consider only the part of the spin-orbit interaction Hamiltonian that is independent of $\mathbf{k}$ because the contribution of the $\mathbf{k}$ dependent part is much smaller. Solving this equation for $\mathbf{k} = 0$ (Γ point) yields the Bloch factors u_{j0} which form a complete and orthonormal basis. The Bloch factor $u_{n\mathbf{k}}$ is expanded for any value of $\mathbf{k}$ using the known Bloch factors u_{j0} at the Γ point

$$u_{n\mathbf{k}}(\mathbf{x}) = \sum_{j=1}^{8} a_j(\mathbf{k})\, u_{j0}(\mathbf{x}). \tag{3.7}$$

For our $\mathbf{k}\cdot\mathbf{p}$ model, the index j goes from 1 to 8 for 8×8 $\mathbf{k}\cdot\mathbf{p}$ (one conduction and three valence bands, including spin), and from 1 to 6 for 6×6 $\mathbf{k}\cdot\mathbf{p}$ (three valence bands, including spin). However, in our algorithmic implementation for 6×6 $\mathbf{k}\cdot\mathbf{p}$ it goes from 3 to 8 because in this case we use the same Hamiltonian matrix (eq. (3.10)) and omit the indices 1 and 2 related to the conduction band.

The band structure near the Γ point is described by perturbation theory around $\mathbf{k} = 0$ using a number of perturbationally defined parameters. The Γ point electron wave function is expanded into s and p orbital functions. A perturbation model that includes the spin-orbit interaction which is responsible for the splitting Δ_{so} between the Γ_7 and Γ_8 valence bands, requires a basis of eight so-called Bloch functions

$$\{|S\uparrow\rangle, |S\downarrow\rangle, |X\uparrow\rangle, |Y\uparrow\rangle, |Z\uparrow\rangle, |X\downarrow\rangle, |Y\downarrow\rangle, |Z\downarrow\rangle\}, \tag{3.8}$$

where X, Y, Z are the p-type Bloch functions referring to the three principal directions in the crystal and the arrows denote the spin. The designations S, X, Y, Z refer to the corresponding symmetry properties under operations of the tetrahedral group. For heterostructures, the envelope functions ψ that correspond to the Bloch functions in eq. (3.8) are given by

$$\{\psi_{S\uparrow}, \psi_{S\downarrow}, \psi_{X\uparrow}, \psi_{Y\uparrow}, \psi_{Z\uparrow}, \psi_{X\downarrow}, \psi_{Y\downarrow}, \psi_{Z\downarrow}\}. \tag{3.9}$$

The structure of the bulk 8×8 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian operator $\mathbf{H}_0$ without strain and

without spin-orbit coupling in the basis of eq. (3.8) is given by

	$\|S\uparrow\rangle$	$\|S\downarrow\rangle$	$\|X\uparrow\rangle$	$\|Y\uparrow\rangle$	$\|Z\uparrow\rangle$	$\|X\downarrow\rangle$	$\|Y\downarrow\rangle$	$\|Z\downarrow\rangle$
$\|S\uparrow\rangle$	H_{cc}	0	$\mathbf{H}_{\mathrm{cv}}$			$\mathbf{0}$		
$\|S\downarrow\rangle$	0	H_{cc}	$\mathbf{0}$			$\mathbf{H}_{\mathrm{cv}}$		
$\|X\uparrow\rangle$								
$\|Y\uparrow\rangle$	$\mathbf{H}_{\mathrm{vc}}$	$\mathbf{0}$	$\mathbf{H}_{\mathrm{vv}}$			$\mathbf{0}$		
$\|Z\uparrow\rangle$								
$\|X\downarrow\rangle$								
$\|Y\downarrow\rangle$	$\mathbf{0}$	$\mathbf{H}_{\mathrm{vc}}$	$\mathbf{0}$			$\mathbf{H}_{\mathrm{vv}}$		
$\|Z\downarrow\rangle$								

. (3.10)

It describes the electrons in the Γ_6 conduction band, or the Γ_7 or Γ_8 valence bands. Our choice of ordering is due to the fact that we are using the same routines within our algorithm for the setup of the 8×8 and the 6×6 Hamiltonian. In the latter case, the first two rows and the first two columns are ignored. For zinc blende, $\mathbf{H}_{\mathrm{vv}}$ is given by

$$\mathbf{H}_{\mathrm{vv}} = \begin{pmatrix} E_{\mathrm{v,av}} + \frac{\hbar^2}{2m_0}\mathbf{k}^2 & & \\ & E_{\mathrm{v,av}} + \frac{\hbar^2}{2m_0}\mathbf{k}^2 & \\ & & E_{\mathrm{v,av}} + \frac{\hbar^2}{2m_0}\mathbf{k}^2 \end{pmatrix} + \tag{3.11}$$

$$\begin{pmatrix} k_xLk_x + k_yMk_y + k_zMk_z & k_xN^+k_y + k_yN^-k_x & k_xN^+k_z + k_zN^-k_x \\ k_yN^+k_x + k_xN^-k_y & k_xMk_x + k_yLk_y + k_zMk_z & k_yN^+k_z + k_zN^-k_y \\ k_zN^+k_x + k_xN^-k_z & k_zN^+k_y + k_yN^-k_z & k_xMk_x + k_yMk_y + k_zLk_z \end{pmatrix},$$

where $E_{\mathrm{v,av}}$ is the energy of the average of the three valence band edges, without strain shifts and without taking spin-orbit splitting into account (see Fig. 3.1). The contribution of the free electron term

$$\frac{\hbar^2}{2m_0}\mathbf{k}^2 = \frac{\hbar^2}{2m_0}\left(k_x^2 + k_y^2 + k_z^2\right) \tag{3.12}$$

could in principle be incorporated into the L and M parameters. This will be discussed further below. For wurtzite the second term has to be replaced by

$$\begin{pmatrix} k_xL_1k_x + k_yM_1k_y + k_zM_2k_z & k_xN_1^+k_y + k_yN_1^-k_x & k_xN_2^+k_z + k_zN_2^-k_x \\ k_yN_1^+k_x + k_xN_1^-k_y & k_xM_1k_x + k_yL_1k_y + k_zM_2k_z & k_yN_2^+k_z + k_zN_2^-k_y \\ k_zN_2^+k_x + k_xN_2^-k_z & k_zN_2^+k_y + k_yN_2^-k_z & k_xM_3k_x + k_yM_3k_y + k_zL_2k_z \end{pmatrix}. \tag{3.13}$$

The DKK (Dresselhaus–Kip–Kittel) parameters [DKK55] L, M, N^+, N^- can either be calculated from the appropriate matrix elements (see Refs. [LYVW09, Hac02]) or

from the Luttinger parameters γ_1, γ_2, γ_3, κ [Bah90, And09]

$$L = F + 2G = (-\gamma_1 - 4\gamma_2 - 1)\,\frac{\hbar^2}{2m_0} \tag{3.14}$$

$$M = H_1 + H_2 = (2\gamma_2 - \gamma_1 - 1)\,\frac{\hbar^2}{2m_0} \tag{3.15}$$

$$N^+ = F - G = (-3\gamma_3 - (3\kappa+1))\frac{\hbar^2}{2m_0} = \frac{N}{2} - (3\kappa+1)\frac{\hbar^2}{2m_0} \tag{3.16}$$

$$N^- = H_1 - H_2 = (-3\gamma_3 + (3\kappa+1))\frac{\hbar^2}{2m_0} = \frac{N}{2} + (3\kappa+1)\frac{\hbar^2}{2m_0}. \tag{3.17}$$

As k_i and k_j commute in bulk, it holds

$$N = N^+ + N^- = F - G + H_1 - H_2 = -6\gamma_3\frac{\hbar^2}{2m_0}. \tag{3.18}$$

The inverse relations for F, G, H_1 and H_2 are

$$F = \frac{1}{3}(L + 2N^+) = \left(-\frac{1}{3}\gamma_1 - \frac{4}{3}\gamma_2 - 2\gamma_3 - 2\kappa - 1\right)\frac{\hbar^2}{2m_0} = -6\sigma\frac{\hbar^2}{2m_0} \tag{3.19}$$

$$G = \frac{1}{3}(L - N^+) = \left(-\frac{1}{3}\gamma_1 - \frac{4}{3}\gamma_2 + \gamma_3 + \kappa\right)\frac{\hbar^2}{2m_0} = -6\delta\frac{\hbar^2}{2m_0} \tag{3.20}$$

$$H_1 = \frac{1}{2}(M + N^-) = \left(-\frac{1}{2}\gamma_1 + \gamma_2 - \frac{3}{2}\gamma_3 + \frac{3}{2}\kappa\right)\frac{\hbar^2}{2m_0} = -6\pi\frac{\hbar^2}{2m_0} \tag{3.21}$$

$$H_2 = \frac{1}{2}(M - N^-) = \left(-\frac{1}{2}\gamma_1 + \gamma_2 + \frac{3}{2}\gamma_3 - \frac{3}{2}\kappa - 1\right)\frac{\hbar^2}{2m_0}, \tag{3.22}$$

where the Foreman parameters σ, π and δ will be introduced further below. Rather than specifying the four parameters L, M, N^+, N^-, occasionally another set of parameters L, M, N, K is specified, where $K = -\frac{\hbar^2}{2m_0}2(3\kappa+1)$ [Lut56]. The parameters F, G, H_1 and H_2 are defined in Ref. [Law71]. There, also an additional fifth Luttinger parameter q [Lut56] related to spin-orbit splitting is given which is typically neglected, and also neglected in our work. The inverse relations for the Luttinger parameters are

$$\begin{aligned}
\gamma_1 &= -\frac{1}{3}\,(L + 2M)\,\frac{2m_0}{\hbar^2} - 1 \\
\gamma_2 &= -\frac{1}{6}\,(L - M)\,\frac{2m_0}{\hbar^2} \\
\gamma_3 &= -\frac{1}{6}(N^+ + N^-)\frac{2m_0}{\hbar^2} = -\frac{1}{6}N\frac{2m_0}{\hbar^2} \\
\kappa &= -\frac{1}{6}(N^+ - N^-)\frac{2m_0}{\hbar^2} - \frac{1}{3},
\end{aligned} \tag{3.23}$$

and

$$\begin{aligned}
\gamma_1 &= -\frac{1}{3}\left(F + 2G + 2H_1 + 2H_2\right)\frac{2m_0}{\hbar^2} - 1 \\
\gamma_2 &= -\frac{1}{6}\left(F + 2G - H_1 - H_2\right)\frac{2m_0}{\hbar^2} \\
\gamma_3 &= -\frac{1}{6}\left(F - G + H_1 - H_2\right)\frac{2m_0}{\hbar^2} \\
\kappa &= -\frac{1}{6}\left(F - G - H_1 + H_2\right)\frac{2m_0}{\hbar^2} - \frac{1}{3}.
\end{aligned} \tag{3.24}$$

The parameter H_2 is small and thus it is often neglected, e.g. in Ref. [For97]. This leads to $N^- \approx M$ and $N^+ = N - N^- \approx N - M$. This is exactly equivalent to the case where κ is not known, and therefore approximated by

$$\kappa \approx -\frac{1}{6}\left(N - 2M\right)\frac{2m_0}{\hbar^2} - \frac{1}{3} = -\frac{1}{3}\left(\gamma_1 - 2\gamma_2 - 3\gamma_3 + 2\right), \tag{3.25}$$

where N is defined in eq. (3.18). Using this definition for κ, i.e. assuming $H_2 = 0$, eq. (3.16) and eq. (3.17) can be expressed using the Luttinger parameters

$$N^- \approx H_1 = M = (2\gamma_2 - \gamma_1 - 1)\frac{\hbar^2}{2m_0} \tag{3.26}$$

$$N^+ = F - G = N - N^- \approx N - M = (-6\gamma_3 - (2\gamma_2 - \gamma_1 - 1))\frac{\hbar^2}{2m_0}. \tag{3.27}$$

For the $\mathbf{k}\cdot\mathbf{p}$ dispersion of bulk semiconductors without magnetic field, the contribution of the term $3\kappa + 1$ to N^+ (eq. (3.16)) and N^- (eq. (3.17)) effectively cancel each other. Therefore, κ is not needed and can be ignored, suggesting to use only the parameter N for the bulk Hamiltonian. This misleads to effectively using $N^+ = N^- = N/2$ (eq. (3.18)), a practice that was adopted by the whole $\mathbf{k}\cdot\mathbf{p}$ community until the last decade. However, Foreman identified this symmetrized $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian to be incorrect for heterostructures [For93], pointing out the noncommutativity of the momentum and position operators in heterostructures. It is thus crucial to use the correct form of the nonsymmetrized $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian which includes the correct definitions of N^+ and N^-. In both cases the entire matrix is Hermitian whereas in the symmetrized approach, additionally, each matrix element is Hermitian. The symmetrized Hamiltonian has been derived from the bulk $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, whereas the nonsymmetrized version is based on Burt's exact envelope function theory for heterostructures [Bur92], which has been extended by Foreman to multi-band $\mathbf{k}\cdot\mathbf{p}$. For that reason it is usually called the Burt–Foreman Hamiltonian [LYVW09]. If the $\mathbf{k}\cdot\mathbf{p}$ material parameters do not depend on position, e.g. in the case of a quantum well with infinite barriers, i.e. no material interfaces, both symmetrizations lead to the same results in the case of zero magnetic field. If one is only interested in the bulk $\mathbf{k}\cdot\mathbf{p}$ dispersion, there is no need to explicitly use N^+ and N^-, and thus N can be used instead (eq. (3.18)). Unfortunately, in a lot of articles in the last decades N^+ and N^- (and also κ) have been ignored and only N

has been used, which we now know is definitely not correct for heterostructures. The noncommutativity of the off-diagonal matrix elements has already been pointed out in the early work of Luttinger [Lut56], where he derived the most general form of the 6×6 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian in the presence of an external homogeneous magnetic field.

Foreman introduced another set of dimensionless parameters σ, π and δ [For93]

$$\begin{aligned}
\sigma &= -\frac{1}{6}F\frac{2m_0}{\hbar^2} = -\frac{1}{18}(L+2N^+)\frac{2m_0}{\hbar^2} \approx -\frac{1}{2}\delta+\overline{\gamma} \\
\pi &= -\frac{1}{6}H_1\frac{2m_0}{\hbar^2} = -\frac{1}{12}(M+N^-)\frac{2m_0}{\hbar^2} \approx \frac{3}{2}\delta+\mu \\
\delta &= -\frac{1}{6}G\frac{2m_0}{\hbar^2} = -\frac{1}{18}(L-N^+)\frac{2m_0}{\hbar^2} \approx \frac{1}{9}(\gamma_1+\gamma_2-3\gamma_3+1),
\end{aligned} \tag{3.28}$$

where $\overline{\gamma}$ and μ are defined as

$$\overline{\gamma} = \frac{1}{2}(\gamma_3+\gamma_2) \tag{3.29}$$

$$\mu = \frac{1}{2}(\gamma_3-\gamma_2). \tag{3.30}$$

The notation in Greek letters σ, π and δ is derived from the s, p, d (and f) orbitals of the constituent atoms. Here, the contribution of the f orbitals is neglected which is equivalent to setting $H_2=0$, i.e. approximating κ (eq. (3.25)). The inverse relations show how the Luttinger parameters can be expanded to reflect the symmetry of the interaction of the bands [LYVW09]

$$\begin{aligned}
\gamma_1 &\approx -\frac{1}{3}\left(F+2G+2H_1\right)\frac{2m_0}{\hbar^2} - 1 = 2\sigma+4\pi+4\delta-1 \\
\gamma_2 &\approx -\frac{1}{6}\left(F+2G-H_1\right)\frac{2m_0}{\hbar^2} = \sigma-\pi+2\delta \\
\gamma_3 &\approx -\frac{1}{6}\left(F-G+H_1\right)\frac{2m_0}{\hbar^2} = \sigma+\pi-\delta \\
\kappa &\approx -\frac{1}{6}\left(F-G-H_1\right)\frac{2m_0}{\hbar^2} - \frac{1}{3} = \sigma-\pi-\delta-\frac{1}{3}.
\end{aligned} \tag{3.31}$$

They are similar to eq. (3.24), with the exception that the term H_2 has been neglected. Consequently, the forth parameter κ is not an independent parameter here. It depends on the choice of σ, π and δ, or γ_1, γ_2 and γ_3, respectively. The corresponding relation for κ in terms of the Luttinger parameters is given in eq. (3.25). Finally, we list the related equations for the DKK parameters

$$\begin{aligned}
L = F+2G &= \left(-6\sigma-12\delta\right)\frac{\hbar^2}{2m_0} \\
N^+ = F-G &= \left(-6\sigma+6\delta\right)\frac{\hbar^2}{2m_0} \\
M \approx H_1 &= -6\pi\frac{\hbar^2}{2m_0} \\
N^- \approx H_1 &= -6\pi\frac{\hbar^2}{2m_0}.
\end{aligned} \tag{3.32}$$

The Luttinger parameters are based on the method of invariants whereas the DKK parameters are based on the perturbation method for deriving the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian matrix. This is the reason for the various definitions of $\mathbf{k}\cdot\mathbf{p}$ parameters.

For the DKK parameters L, M, N, there is another frequently used definition in the literature (e.g. used by Bir and Pikus [BP74] and in Ref. [VSW07]), also called L, M, N, which often causes confusion in the numerical values of the parameters and thus even in some cases leading to incorrect use of parameters (see Section 3.4). They originally were used by Luttinger and Kohn [LK55]. There, they were termed A, B, C, and should not be confused with A, B, C of eq. (3.137), eq. (3.138) and eq. (3.139). These alternative LK (Luttinger–Kohn) parameters, labeled with superscript 'LK', read

$$L^{\mathrm{LK}} = A^{\mathrm{LK}} = L + \frac{\hbar^2}{2m_0} = (-\gamma_1 - 4\gamma_2)\,\frac{\hbar^2}{2m_0} \tag{3.33}$$

$$M^{\mathrm{LK}} = B^{\mathrm{LK}} = M + \frac{\hbar^2}{2m_0} = (2\gamma_2 - \gamma_1)\,\frac{\hbar^2}{2m_0} \tag{3.34}$$

$$N^{\mathrm{LK}} = C^{\mathrm{LK}} = N^{\mathrm{LK}+} + N^{\mathrm{LK}-} = N = -6\gamma_3\frac{\hbar^2}{2m_0} \tag{3.35}$$

$$N^{\mathrm{LK}+} = N^{+} \approx N^{\mathrm{LK}} - \left(M^{\mathrm{LK}} - \frac{\hbar^2}{2m_0}\right) = N - M \tag{3.36}$$

$$N^{\mathrm{LK}-} = N^{-} \approx M^{\mathrm{LK}} - \frac{\hbar^2}{2m_0} = M. \tag{3.37}$$

Here, L^{LK} and M^{LK} are defined including the free electron term $\frac{\hbar^2}{2m_0}$ (see eq. (3.12)). Expressing them using the Luttinger parameters now differs because the term '-1' that is present in eq. (3.14) and eq. (3.15) has disappeared. Consequently, the diagonal term $\frac{\hbar^2}{2m_0}\mathbf{k}^2$ must be omitted in $\mathbf{H}_{\mathrm{vv}}$, i.e. the term $E_{\mathrm{v,av}} + \frac{\hbar^2}{2m_0}\mathbf{k}^2$ in eq. (3.11) has to be replaced by $E_{\mathrm{v,av}}$. The inverse relations for the Luttinger parameters read

$$\gamma_1 = -\frac{1}{3}\left(L^{\mathrm{LK}} + 2M^{\mathrm{LK}}\right)\frac{2m_0}{\hbar^2} = -\frac{1}{3}\left(L + 2M\right)\frac{2m_0}{\hbar^2} - 1 \tag{3.38}$$

$$\gamma_2 = -\frac{1}{6}\left(L^{\mathrm{LK}} - M^{\mathrm{LK}}\right)\frac{2m_0}{\hbar^2} = -\frac{1}{6}\left(L - M\right)\frac{2m_0}{\hbar^2} \tag{3.39}$$

$$\gamma_3 = -\frac{1}{6}N^{\mathrm{LK}}\frac{2m_0}{\hbar^2} = -\frac{1}{6}N\frac{2m_0}{\hbar^2} \tag{3.40}$$

$$\kappa = -\frac{1}{6}(N^{\mathrm{LK}+} - N^{\mathrm{LK}-})\frac{2m_0}{\hbar^2} - \frac{1}{3} = -\frac{1}{6}(N^{+} - N^{-})\frac{2m_0}{\hbar^2} - \frac{1}{3}, \tag{3.41}$$

where the formula for γ_1 differs for the two possible definitions of the DKK or LK parameters. The term '-1' in eq. (3.38) shows how they are related to each other. The relations for γ_2, γ_3, and κ correspond to the ones given in eq. (3.23). In order to avoid confusion, we recommend to provide values for the Luttinger parameters when publishing or comparing material parameters, rather than the ambiguous L, M, N parameters.

For wurtzite the Rashba-Sheka-Pikus (RSP) parameters of the valence band A_1, A_2, A_3, A_4, A_5, A_6 are similar to the Luttinger parameters in zinc blende. The A_7 parameter

that describes the $\mathbf{k}$ dependent spin-orbit splitting is usually neglected. From those six parameters one can obtain the nine parameters L_1, L_2, M_1, M_2, M_3, N_1^+, N_1^-, N_2^+, N_2^- as follows [FB03]

$$\begin{aligned}
L_1 &= \frac{\hbar^2}{2m_0}\left(A_5+\tilde{A}\right) = -6(\sigma+\delta) \qquad (3.42)\\
L_2 &= \frac{\hbar^2}{2m_0}\left(A_1-1\right) = -6\sigma_z\\
M_1 &= \frac{\hbar^2}{2m_0}\left(-A_5+\tilde{A}\right) = -6\delta\\
M_2 &= \frac{\hbar^2}{2m_0}\hat{A} = -6\pi\\
M_3 &= \frac{\hbar^2}{2m_0}\left(A_2-1\right) = -6\pi_z\\
N_1^+ &= \frac{\hbar^2}{2m_0}\left(3A_5-\tilde{A}\right) = N_1 - M_1 = -6(\sigma-\delta)\\
N_1^- &= \frac{\hbar^2}{2m_0}\left(-A_5+\tilde{A}\right) = M_1 = -6\delta\\
N_2^+ &= \frac{\hbar^2}{2m_0}\left(\sqrt{2}A_6-\hat{A}\right) = N_2 - M_2 = -6\sigma_{xz}\\
N_2^- &= \frac{\hbar^2}{2m_0}\hat{A} = M_2 = -6\pi,
\end{aligned}$$

where we used

$$\tilde{A} = A_2 + A_4 - 1 \qquad (3.43)$$

$$\hat{A} = A_1 + A_3 - 1. \qquad (3.44)$$

It also holds

$$N_1 = N_1^+ + N_1^- = L_1 - M_1 = \frac{\hbar^2}{2m_0}2A_5 = -6\sigma \qquad (3.45)$$

$$N_2 = N_2^+ + N_2^- = \frac{\hbar^2}{2m_0}\sqrt{2}A_6 = -6\sigma_{xz} - 6\pi. \qquad (3.46)$$

The 'Foreman' parameters σ, σ_z, σ_{xz}, π, π_z and δ have actually been introduced by Mireles and Ulloa [MU99]. The relation $N_1 = L_1 - M_1$ is due to the sixfold rotational symmetry of the Hamiltonian [CC96]. The related matrix elements are given in Refs. [LYVW09, MU99]. The contribution of the term '-1' (free electron term) for L_1, L_2, M_1, M_2 and M_3 has the same origin as in the zinc blende case. Thus one should be careful when comparing material parameters because for wurtzite there are also two

definitions possible. The inverse relations to eq. (3.42) read [CC96, MU99]

$$A_1 = \frac{2m_0}{\hbar^2}L_2 + 1 = 1 - 6\sigma_z \tag{3.47}$$

$$A_2 = \frac{2m_0}{\hbar^2}M_3 + 1 = 1 - 6\pi_z \tag{3.48}$$

$$A_3 = \frac{2m_0}{\hbar^2}\left(M_2 - L_2\right) = -6(\pi - \sigma_z) \tag{3.49}$$

$$A_4 = \frac{2m_0}{\hbar^2}\frac{1}{2}\left(L_1 + M_1 - 2M_3\right) = -3\sigma - 6(\delta - \pi_z) \tag{3.50}$$

$$A_5 = \frac{2m_0}{\hbar^2}\frac{1}{2}\left(L_1 - M_1\right) = \frac{2m_0}{\hbar^2}\frac{1}{2}N_1 = -3\sigma \tag{3.51}$$

$$A_6 = \frac{2m_0}{\hbar^2}\frac{\sqrt{2}}{2}N_2 = -\frac{6}{\sqrt{2}}(\pi + \delta). \tag{3.52}$$

Finally, the 'Foreman' parameters can be expressed as

$$\sigma = -\frac{1}{6}(L_1 - M_1)\frac{2m_0}{\hbar^2} = -\frac{1}{6}N_1\frac{2m_0}{\hbar^2} = -\frac{1}{3}A_5 \tag{3.53}$$

$$\sigma_z = -\frac{1}{6}L_2\frac{2m_0}{\hbar^2} = -\frac{1}{6}(A_1 - 1) \tag{3.54}$$

$$\begin{aligned}\sigma_{xz} &= -\frac{1}{6}(N_2 - M_2)\frac{2m_0}{\hbar^2} = -\frac{1}{6}N_2^+\frac{2m_0}{\hbar^2} = -\frac{1}{6}\left(-A_1 - A_3 + \sqrt{2}A_6 + 1\right) \\ &= -\frac{1}{6}\left(\sqrt{2}A_6 - \hat{A}\right)\end{aligned} \tag{3.55}$$

$$\pi = -\frac{1}{6}M_2\frac{2m_0}{\hbar^2} = -\frac{1}{6}N_2^-\frac{2m_0}{\hbar^2} = -\frac{1}{6}(A_1 + A_3 - 1) = -\frac{1}{6}\hat{A} \tag{3.56}$$

$$\pi_z = -\frac{1}{6}M_3\frac{2m_0}{\hbar^2} = -\frac{1}{6}(A_2 - 1) \tag{3.57}$$

$$\delta = -\frac{1}{6}M_1\frac{2m_0}{\hbar^2} = -\frac{1}{6}(A_2 + A_4 - A_5 - 1) = -\frac{1}{6}(\tilde{A} - A_5). \tag{3.58}$$

H_{cc} in eq. (3.10) is defined as

$$H_{\mathrm{cc}} = E_{\mathrm{c}} + k_x A_{\mathrm{c}} k_x + k_y A_{\mathrm{c}} k_y + k_z A_{\mathrm{c}} k_z \tag{3.59}$$

for zinc blende, and

$$H_{\mathrm{cc}} = E_{\mathrm{c}} + k_x A_{\mathrm{c}2} k_x + k_y A_{\mathrm{c}2} k_y + k_z A_{\mathrm{c}1} k_z \tag{3.60}$$

for wurtzite, where E_{c} is the conduction band edge (without strain shifts). The parameter A_{c} is defined as $A_{\mathrm{c}} = \frac{\hbar^2}{2m_0}S$ and S is a dimensionless parameter defined for zinc blende as [Hac02]

$$S = 1 + 2F = \left(1 + 2\frac{1}{m_0}\sum_{n\in B}\frac{\left|\langle S\left|p_x\right|n\rangle\right|^2}{E_{\mathrm{c}} - E_n}\right), \tag{3.61}$$

where the summation is over all bands apart from the valence bands. Here, F should not be confused with the parameter F of eq. (3.19). In contrast to our definition of the valence band parameters L, M, N^+ and N^-, the conduction band parameter A_c includes the free electron term $\frac{\hbar^2}{2m_0}$ which corresponds to the '1' in eq. (3.61). If the free electron term were not included, then it must be included explicitly into the H_cc term, analogous to the first line of eq. (3.11) for the valence band part $\mathbf{H}_\mathrm{vv}$. S can also be evaluated through the experimentally determined conduction band mass m_e at the Γ point using the relation [LFC96]

$$S = \frac{m_0}{m_\mathrm{e}} - \frac{2E_\mathrm{P}}{3E_\mathrm{gap}} - \frac{E_\mathrm{P}}{3(E_\mathrm{gap}+\Delta_\mathrm{so})} = \frac{m_0}{m_\mathrm{e}} - E_\mathrm{P}\frac{E_\mathrm{gap}+\frac{2}{3}\Delta_\mathrm{so}}{E_\mathrm{gap}\,(E_\mathrm{gap}+\Delta_\mathrm{so})}, \tag{3.62}$$

where E_gap is the (unstrained) band gap energy between the lowest conduction band edge and the highest valence band edge energy. If one wants to switch off the coupling between electrons and holes, simply setting $E_\mathrm{P} = 0\,\mathrm{eV}$ leads to an isotropic and parabolic energy dispersion for the electrons. In this case, the S parameter is the inverse of the effective electron mass, $S = \frac{m_0}{m_\mathrm{e}}$, giving an intuitive meaning to this $\mathbf{k}\cdot\mathbf{p}$ parameter. In Ref. [VMRM01] the dimensionless F parameter (eq. (3.61)) is given for all zinc blende materials, where $F = (S-1)/2$. However, as the band gap E_gap is temperature dependent (eq. (B.2)), next**nano**[3] by default calculates S directly from the actual band gap and the effective electron mass m_e, rather than using the S parameter of the database. We notice that different definitions of the S or A_c parameter occur in the literature depending on whether the free electron term is included or not. Thus one has to be careful when comparing different sets of material parameters. For wurtzite, the parameter A_{ci} is defined as $A_{ci} = \frac{\hbar^2}{2m_0}S_i$ with index $i = \{1,2\}$. The index $i = 1$ refers to the direction parallel to the hexagonal c axis and $i = 2$ to the directions perpendicular to it. It holds $S_i = 1 + 2F_i$ and [CC96]

$$S_1 = \frac{m_0}{m_{\mathrm{e},\parallel}} - E_\mathrm{P1}\frac{E_\mathrm{gap}+2\Delta_2}{(E_\mathrm{gap}+\Delta_1+\Delta_2)(E_\mathrm{gap}+2\Delta_2)-2\Delta_3^2} \tag{3.63}$$

$$S_2 = \frac{m_0}{m_{\mathrm{e},\perp}} - E_\mathrm{P2}\frac{(E_\mathrm{gap}+\Delta_1+\Delta_2)(E_\mathrm{gap}+\Delta_2)-\Delta_3^2}{E_\mathrm{gap}\left[(E_\mathrm{gap}+\Delta_1+\Delta_2)(E_\mathrm{gap}+2\Delta_2)-2\Delta_3^2\right]}. \tag{3.64}$$

In the limit $\Delta_1 = 0$, $\Delta_2 = \Delta_3 = \frac{1}{3}\Delta_\mathrm{so}$, the zinc blende result (eq. (3.62)) is obtained.

The coupling between conduction and valence bands in eq. (3.10) is defined as

$$\mathbf{H}_\mathrm{cv} = \begin{pmatrix} H_\mathrm{cv}^{SX} & H_\mathrm{cv}^{SY} & H_\mathrm{cv}^{SZ} \end{pmatrix}, \tag{3.65}$$

and

$$\mathbf{H}_\mathrm{vc} = \begin{pmatrix} H_\mathrm{vc}^{SX} \\ H_\mathrm{vc}^{SY} \\ H_\mathrm{vc}^{SZ} \end{pmatrix}. \tag{3.66}$$

For zinc blende, the components are given by

$$H_\mathrm{cv}^{SX} = k_y B k_z + iPk_x \tag{3.67}$$

$$H_\mathrm{cv}^{SY} = k_z B k_x + iPk_y \tag{3.68}$$

$$H_\mathrm{cv}^{SZ} = k_x B k_y + iPk_z, \tag{3.69}$$

and

$$H_{\mathrm{vc}}^{SX} = k_z B k_y - i k_x P \tag{3.70}$$
$$H_{\mathrm{vc}}^{SY} = k_x B k_z - i k_y P \tag{3.71}$$
$$H_{\mathrm{vc}}^{SZ} = k_y B k_x - i k_z P. \tag{3.72}$$

For wurtzite they are

$$H_{\mathrm{cv}}^{SX} = k_y B_1 k_z + i P_2 k_x \tag{3.73}$$
$$H_{\mathrm{cv}}^{SY} = k_z B_2 k_x + i P_2 k_y \tag{3.74}$$
$$H_{\mathrm{cv}}^{SZ} = k_x B_3 k_y + i P_1 k_z, \tag{3.75}$$

and

$$H_{\mathrm{vc}}^{SX} = k_z B_1 k_y - i k_x P_2 \tag{3.76}$$
$$H_{\mathrm{vc}}^{SY} = k_x B_2 k_z - i k_y P_2 \tag{3.77}$$
$$H_{\mathrm{vc}}^{SZ} = k_y B_3 k_x - i k_z P_1. \tag{3.78}$$

We notice that $\mathbf{H}_{\mathrm{vc}}$ is not the Hermitian conjugate of $\mathbf{H}_{\mathrm{cv}}$ because the operator ordering is different. This ordering is the one suggested by Foreman [For97]. The correct form of the inversion asymmetry parameter B has been derived by Loehr [Loe95]. B is zero for materials that possess inversion symmetry as is the case for diamond-type crystals. Zinc blende crystals do not have inversion symmetry, thus $B \neq 0$. This also applies to wurtzite crystals, thus $B_1 \neq 0$, $B_2 \neq 0$ and $B_3 \neq 0$. For one-dimensional simulations, B only contributes for nonzero $\mathbf{k}_{\parallel}$ vectors. It is common practice in the $\mathbf{k}\cdot\mathbf{p}$ literature to neglect the contribution of the B parameter. However, this inversion asymmetry parameter is responsible for a spin-splitting of the bulk $\mathbf{k}\cdot\mathbf{p}$ dispersion for the bands along directions other than [001] and [111], where the splitting is suppressed by symmetry elements of the group of $\mathbf{k}$ [ESC87] (see Fig. 3.11). For heterostructures the B parameter leads to a splitting of the states for any direction of $\mathbf{k}_{\parallel}$, even if the structure itself has an inversion center, i.e. no structural inversion asymmetry (SIA). In Section 3.6, we present an example (Fig. 3.17). The optical matrix parameter P (Kane momentum matrix element) that mixes the conduction and valence band states is given by

$$P = -\frac{i\hbar}{m_0} \langle S | p_x | X \rangle = \sqrt{\frac{\hbar^2}{2m_0} E_{\mathrm{P}}} \tag{3.79}$$
$$E_{\mathrm{P}} = \frac{2m_0}{\hbar^2} P^2, \tag{3.80}$$

where the Kane parameter E_{P} is the energy equivalent to P, whose values are usually

around 22 eV for almost all semiconductors. They are similar for wurtzite

$$P_1 = -\frac{i\hbar}{m}\langle S|p_z|Z\rangle = \sqrt{\frac{\hbar^2}{2m_0}E_{\mathrm{P1}}} \tag{3.81}$$

$$P_2 = -\frac{i\hbar}{m}\langle S|p_x|X\rangle = -\frac{i\hbar}{m}\langle S|p_y|Y\rangle = \sqrt{\frac{\hbar^2}{2m_0}E_{\mathrm{P2}}}. \tag{3.82}$$

In bulk, P only contributes for nonzero $\mathbf{k}$ vectors, whereas for heterostructures it is also relevant for $\mathbf{k}_\parallel = 0$.

Any algorithmic implementation of the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian should be the one of the wurtzite Hamiltonian. It implicitly contains the zinc blende and diamond-type cases by setting $L_1 = L_2 = L$, $M_1 = M_2 = M_3 = M$, $N_1^+ = N_2^+ = N^+$, $N_1^- = N_2^- = N^-$, $P_1 = P_2 = P$, $B_1 = B_2 = B_3 = B$ and $S_1 = S_2 = S$.

Spin-orbit coupling The relativistic effect of spin is approximately taken into account by including an additional term $\mathbf{H}_{\mathrm{so}}$ in the Schrödinger equation (eq. (3.1)). Using the definition

$$\Delta_{\mathrm{so}} = -3i\left(\frac{\hbar^2}{4m_0^2c^2}\right)\left\langle X\left|(\nabla V\times\mathbf{p})_y\right|Z\right\rangle, \tag{3.83}$$

we can write down the spin-orbit interaction Hamiltonian for zinc blende [HS90]. In the basis of eq. (3.8) it reads

$$\mathbf{H}_{\mathrm{so}} = \frac{1}{3}\Delta_{\mathrm{so}}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & -i & 0 & 0 & 0 & 1\\ 0 & 0 & i & 0 & 0 & 0 & 0 & -i\\ 0 & 0 & 0 & 0 & 0 & -1 & i & 0\\ 0 & 0 & 0 & 0 & -1 & 0 & i & 0\\ 0 & 0 & 0 & 0 & -i & -i & 0 & 0\\ 0 & 0 & 1 & i & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3.84}$$

where Δ_{so} is the spin-orbit split-off energy. The spin-orbit Hamiltonian matrix can be diagonalized if one chooses the Bloch basis [LFC96]

$$
\begin{aligned}
|u_{\mathrm{e}}\uparrow\rangle &= \left|\frac{1}{2},\frac{1}{2}\right\rangle_{\mathrm{e}} = |S\uparrow\rangle \\
|u_{\mathrm{e}}\downarrow\rangle &= \left|\frac{1}{2},-\frac{1}{2}\right\rangle_{\mathrm{e}} = |iS\downarrow\rangle \\
|u_{\mathrm{hh}}\uparrow\rangle &= \left|\frac{3}{2},\frac{3}{2}\right\rangle = \frac{1}{\sqrt{2}}\,|(X+iY)\uparrow\rangle \\
|u_{\mathrm{hh}}\downarrow\rangle &= \left|\frac{3}{2},-\frac{3}{2}\right\rangle = \frac{i}{\sqrt{2}}\,|(X-iY)\downarrow\rangle \\
|u_{\mathrm{lh1}}\rangle &= \left|\frac{3}{2},\frac{1}{2}\right\rangle = \frac{i}{\sqrt{6}}\,[|(X+iY)\downarrow\rangle - 2\,|Z\uparrow\rangle] \\
|u_{\mathrm{lh2}}\rangle &= \left|\frac{3}{2},-\frac{1}{2}\right\rangle = \frac{1}{\sqrt{6}}\,[|(X-iY)\uparrow\rangle + 2\,|Z\downarrow\rangle] \\
|u_{\mathrm{so1}}\rangle &= \left|\frac{1}{2},\frac{1}{2}\right\rangle = \frac{1}{\sqrt{3}}\,[|(X+iY)\downarrow\rangle + |Z\uparrow\rangle] \\
|u_{\mathrm{so2}}\rangle &= \left|\frac{1}{2},-\frac{1}{2}\right\rangle = \frac{i}{\sqrt{3}}\,[|-(X-iY)\uparrow\rangle + |Z\downarrow\rangle]\,,
\end{aligned}
\tag{3.85}
$$

leading to the eigenstates known as heavy hole (hh), light hole (lh) and spin-orbit split-off hole (so) with positive and negative angular momentum projection. The prefactors are normalization constants and these linear combinations are known as the 'angular momentum representation'. The basis states for the electrons are included for completeness. They are not affected by the spin-orbit interaction. The heavy and light holes are degenerate and their eigenvalues are $\frac{1}{3}\Delta_{\mathrm{so}}$, whereas the eigenvalue of the split-off hole is $-\frac{2}{3}\Delta_{\mathrm{so}}$). Here we classified the six valence states in terms of $|\mathbf{J}, J_z\rangle$ states with the use of the Clebsch-Gordan coefficients for the angular momentum $\mathbf{J} = \mathbf{L}_{\mathrm{B}} + \mathbf{S}$ and its z component [EO56]. $\mathbf{L}_{\mathrm{B}}$ is the angular momentum of the Bloch orbit and $\mathbf{S}$ is the spin. The terms 'heavy' and 'light' originate from the distinct curvatures of the energy dispersions of these bands. For nonzero $\mathbf{k}$ values, the states cannot be label any more as pure heavy, light and split-off holes. Projecting the calculated solution in the basis of eq. (3.8) onto the basis given in eq. (3.85), determines if the corresponding spinors are dominated by the character of a heavy, light or split-off hole. Further details about the bulk band dispersion can be found for instance in Ref. [CC92]. As an example we provide the band structure of InAs in Fig. 3.11.

For wurtzite, the spin-orbit interaction Hamiltonian in the basis of eq. (3.8) is given

by

$$\mathbf{H}_{\mathrm{so}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i\Delta_2 & 0 & 0 & 0 & \Delta_3 \\ 0 & 0 & i\Delta_2 & 0 & 0 & 0 & 0 & -i\Delta_3 \\ 0 & 0 & 0 & 0 & 0 & -\Delta_3 & i\Delta_3 & 0 \\ 0 & 0 & 0 & 0 & -\Delta_3 & 0 & i\Delta_2 & 0 \\ 0 & 0 & 0 & 0 & -i\Delta_3 & -i\Delta_2 & 0 & 0 \\ 0 & 0 & \Delta_3 & i\Delta_3 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3.86}$$

with Δ_2 corresponding to the direction parallel to the hexagonal c axis, and Δ_3 to the plane perpendicular to it. It is usually assumed that $\Delta_2 = \Delta_3 = \frac{1}{3}\Delta_{\mathrm{so}}$, which actually yields the zinc blende spin-orbit Hamiltonian (eq. (3.84)).

Crystal field splitting In wurtzite we additionally have to consider the crystal field splitting, i.e. the energy splitting produced by the anisotropy of the hexagonal symmetry. In the basis of eq. (3.8) it is given by

$$\mathbf{H}_{\mathrm{cr}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3.87}$$

where $\Delta_1 = \Delta_{\mathrm{cr}}$ is the crystal field splitting energy which is zero in both diamond and zinc blende materials.

Within the algorithm of the next**nano** program, it is sufficient to only implement the wurtzite spin-orbit Hamiltonian. It includes both crystal field splitting and spin-orbit coupling and reads

$$\mathbf{H}_{\mathrm{so,cr}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta_1 & -i\Delta_2 & 0 & 0 & 0 & \Delta_3 \\ 0 & 0 & i\Delta_2 & \Delta_1 & 0 & 0 & 0 & -i\Delta_3 \\ 0 & 0 & 0 & 0 & 0 & -\Delta_3 & i\Delta_3 & 0 \\ 0 & 0 & 0 & 0 & -\Delta_3 & \Delta_1 & i\Delta_2 & 0 \\ 0 & 0 & 0 & 0 & -i\Delta_3 & -i\Delta_2 & \Delta_1 & 0 \\ 0 & 0 & \Delta_3 & i\Delta_3 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.88}$$

It implicitly includes the zinc blende case. The basis states that diagonalize this wurtzite spin-orbit Hamiltonian are not the same as the ones we used before in the zinc blende

case (eq. (3.85)). They are [CC96]

$$
\begin{aligned}
&|iS\uparrow\rangle \\
&|iS\downarrow\rangle \\
&|u_1\uparrow\rangle = -\frac{1}{\sqrt{2}}\,|(X+iY)\uparrow\rangle \\
&|u_2\uparrow\rangle = \frac{1}{\sqrt{2}}\,|(X-iY)\uparrow\rangle \\
&|u_3\uparrow\rangle = |Z\uparrow\rangle \\
&|u_4\downarrow\rangle = \frac{1}{\sqrt{2}}\,|(X-iY)\downarrow\rangle \\
&|u_5\downarrow\rangle = -\frac{1}{\sqrt{2}}\,|(X+iY)\downarrow\rangle \\
&|u_6\downarrow\rangle = |Z\downarrow\rangle\,.
\end{aligned}
\tag{3.89}
$$

After diagonalization, one obtains the following eigenvalues [CC96]

$$E_1 = E_{\mathrm{A}} = \Delta_1 + \Delta_2 \tag{3.90}$$

$$E_2 = E_{\mathrm{B}} = \frac{\Delta_1 - \Delta_2}{2} + \sqrt{\left(\frac{\Delta_1 - \Delta_2}{2}\right)^2 + 2\Delta_3^2} \tag{3.91}$$

$$E_3 = E_{\mathrm{C}} = \frac{\Delta_1 - \Delta_2}{2} - \sqrt{\left(\frac{\Delta_1 - \Delta_2}{2}\right)^2 + 2\Delta_3^2}, \tag{3.92}$$

where A, B and C refer to the heavy hole, light hole and crystal field split-off hole energies in wurtzite. For zinc blende one has to set $\Delta_1 = 0$ and $\Delta_2 = \Delta_3 = \frac{1}{3}\Delta_{\mathrm{so}}$. The zinc blende eigenvalues can further be simplified to

$$E_1 = E_2 = E_{\mathrm{hh}} = E_{\mathrm{lh}} = \frac{1}{3}\Delta_{\mathrm{so}} \tag{3.93}$$

$$E_3 = E_{\mathrm{so}} = -\frac{2}{3}\Delta_{\mathrm{so}}, \tag{3.94}$$

where heavy hole (hh) and light hole (lh) are degenerate. They are separated from the split-off hole (so) by the spin-orbit splitting energy Δ_{so}. Thus the valence band edges at Γ and the conduction band edges at Γ, L and X are determined as follows

$$E_{\mathrm{c}}^{\Gamma} = E_{\mathrm{v,av}} + \max(E_1, E_2) + E_{\mathrm{gap}}^{\Gamma} \tag{3.95}$$

$$E_{\mathrm{c}}^{L} = E_{\mathrm{v,av}} + \max(E_1, E_2) + E_{\mathrm{gap}}^{L} \tag{3.96}$$

$$E_{\mathrm{c}}^{X} = E_{\mathrm{v,av}} + \max(E_1, E_2) + E_{\mathrm{gap}}^{X} \tag{3.97}$$

$$E_{\mathrm{hh}} = E_{\mathrm{v,av}} + E_1 \tag{3.98}$$

$$E_{\mathrm{lh}} = E_{\mathrm{v,av}} + E_2 \tag{3.99}$$

$$E_{\mathrm{so}} = E_{\mathrm{v,av}} + E_3, \tag{3.100}$$

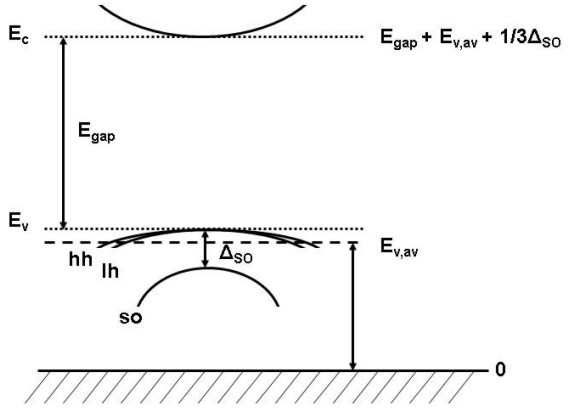


Figure 3.1.: Conduction (E_c) and valence band (E_v) alignment in a zinc blende semiconductor. The heavy hole (hh), light hole (lh) and split-off hole (so) band edges are obtained by adding the spin-orbit Hamiltonian to the average valence band edge energy $E_\mathrm{v,av}$.

where we use the average of the three valence bands $E_\mathrm{v,av}$ as our reference point (Fig. 3.1). In fact, this reference is the average valence band edge energy in the absence of spin-orbit (and crystal field) splitting. This definition is valid for both zinc blende and wurtzite and is used to specify the valence band offset between different materials on a global scale [VdW89]. Very often, however, the valence band offset is instead defined as the difference in energy with respect to the highest hole band edges between two materials.

Modified $\mathbf{k} \cdot \mathbf{p}$ parameters In a 6×6 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian all conduction bands are considered as a perturbation. In 8×8 $\mathbf{k} \cdot \mathbf{p}$ theory, the lowest conduction band is now included in the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian and not treated as a perturbation any more. Thus the related material parameters are different in 6×6 $\mathbf{k} \cdot \mathbf{p}$ and 8×8 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonians. Therefore, the L, M, N^+, N^- parameters in eq. (3.11) must be replaced by the modified DKK parameters L', M', $N^{+\prime}$, $N^{-\prime}$ and N' because the latter include the now required correction term

$$L' = L + \frac{\hbar^2}{2m_0}\frac{E_\mathrm{P}}{E_\mathrm{gap}} = L + \frac{P^2}{E_\mathrm{gap}} \tag{3.101}$$

$$M' = M \tag{3.102}$$

$$N^{+\prime} = N^+ + \frac{\hbar^2}{2m_0}\frac{E_\mathrm{P}}{E_\mathrm{gap}} = N^+ + \frac{P^2}{E_\mathrm{gap}} \tag{3.103}$$

$$N^{-\prime} = N^- \tag{3.104}$$

$$N' = N + \frac{\hbar^2}{2m_0}\frac{E_\mathrm{P}}{E_\mathrm{gap}} = N + \frac{P^2}{E_\mathrm{gap}}. \tag{3.105}$$

The correction term is temperature dependent because obviously the band gap depends on temperature (see Appendix B.2 and Appendix B.3).

The modifications of the L_1, L_2, N_1^+, N_2^+, N_1 and N_2 parameters for wurtzite have to be done in a similar manner [AO00], whereas the M_1, M_2, M_3, N_1^-, N_2^- parameters remain unchanged as in the case of zinc blende

$$L_1' = L_1 + \frac{P_1^2}{E_{\text{gap}}} \tag{3.106}$$

$$L_2' = L_2 + \frac{P_2^2}{E_{\text{gap}}} \tag{3.107}$$

$$N_1^{+\prime} = N_1^+ + \frac{P_1^2}{E_{\text{gap}}} \tag{3.108}$$

$$N_2^{+\prime} = N_2^+ + \frac{P_1 P_2}{E_{\text{gap}}} \tag{3.109}$$

$$N_1' = N_1 + \frac{P_1^2}{E_{\text{gap}}} \tag{3.110}$$

$$N_2' = N_2 + \frac{P_1 P_2}{E_{\text{gap}}}. \tag{3.111}$$

For the RSP parameters the modifications read [AO00]

$$A_1' = A_1 + \frac{E_{\text{P2}}}{E_{\text{gap}}} \tag{3.112}$$

$$A_2' = A_2 \tag{3.113}$$

$$A_3' = A_3 - \frac{E_{\text{P2}}}{E_{\text{gap}}} \tag{3.114}$$

$$A_4' = A_4 + \frac{1}{2}\frac{E_{\text{P1}}}{E_{\text{gap}}} \tag{3.115}$$

$$A_5' = A_5 + \frac{1}{2}\frac{E_{\text{P1}}}{E_{\text{gap}}} \tag{3.116}$$

$$A_6' = A_6 + \frac{\sqrt{2}}{2}\frac{\sqrt{E_{\text{P1}}E_{\text{P2}}}}{E_{\text{gap}}}, \tag{3.117}$$

where E_{P1} refers to the orientation parallel and E_{P2} perpendicular to the hexagonal c axis.

For zinc blende, it holds for the modified Luttinger parameters [PB66]

$$\gamma_1' = \gamma_1 - \frac{1}{3}\frac{E_{\text{P}}}{E_{\text{gap}}} \tag{3.118}$$

$$\gamma_2' = \gamma_2 - \frac{1}{6}\frac{E_{\text{P}}}{E_{\text{gap}}} \tag{3.119}$$

$$\gamma_3' = \gamma_3 - \frac{1}{6}\frac{E_{\text{P}}}{E_{\text{gap}}} \tag{3.120}$$

$$\kappa' = \kappa - \frac{1}{6}\frac{E_{\text{P}}}{E_{\text{gap}}}. \tag{3.121}$$

Finally, we list the modifications of the F, G, H_1, H_2 parameters

$$F' = F + \frac{\hbar^2}{2m_0}\frac{E_\mathrm{P}}{E_\mathrm{gap}} = F + \frac{P^2}{E_\mathrm{gap}} \tag{3.122}$$

$$G' = G \tag{3.123}$$

$$H_1' = H_1 \tag{3.124}$$

$$H_2' = H_2, \tag{3.125}$$

and Foreman's σ, π and δ parameters

$$\sigma' = \sigma - \frac{1}{6}\frac{E_\mathrm{P}}{E_\mathrm{gap}} \tag{3.126}$$

$$\pi' = \pi \tag{3.127}$$

$$\delta' = \delta. \tag{3.128}$$

For wurtzite, the latter are modified to

$$\sigma' = \sigma - \frac{1}{6}\frac{E_\mathrm{P1}}{E_\mathrm{gap}} \tag{3.129}$$

$$\sigma_z' = \sigma_z - \frac{1}{6}\frac{E_\mathrm{P2}}{E_\mathrm{gap}} \tag{3.130}$$

$$\sigma_{xz}' = \sigma_{xz} - \frac{1}{6}\frac{\sqrt{E_\mathrm{P1}E_\mathrm{P2}}}{E_\mathrm{gap}} \tag{3.131}$$

$$\pi' = \pi \tag{3.132}$$

$$\pi_z' = \pi_z \tag{3.133}$$

$$\delta' = \delta. \tag{3.134}$$

I have implemented full flexibility for the user into the next**nano**^{3} software with respect to choice of $\mathbf{k}\cdot\mathbf{p}$ parameters. The user can either specify the 6×6 $\mathbf{k}\cdot\mathbf{p}$ L, M, N parameters or the Luttinger parameters γ_1, γ_2, γ_3. The user can decide whether he wants to include κ or if he wants to approximate κ. The L', M', N' parameters for 8×8 $\mathbf{k}\cdot\mathbf{p}$ can be specified directly, or calculated automatically from the 6×6 $\mathbf{k}\cdot\mathbf{p}$ DKK or Luttinger parameters taking into account the temperature dependent band gap. Additionally, the modified Luttinger parameters for 8×8 $\mathbf{k}\cdot\mathbf{p}$ (with or without κ') can be entered instead. The user can specify the S parameter, or the program calculates S from the temperature dependent band gap and the effective electron mass. By default, a rescaling of the $\mathbf{k}\cdot\mathbf{p}$ parameters is not performed by next**nano**^{3}, in contrast to next**nano**++. However, the user can choose to rescale the parameters automatically to $S=0$ or $S=1$ (eq. (3.158) or eq. (3.158), respectively). This is sometimes necessary in order to avoid spurious solutions (see Section 3.2). In any case, a consistent set of all $\mathbf{k}\cdot\mathbf{p}$ parameters (DKK, Luttinger, Foreman, ...) are written out, also the ones not specified, including the 6×6 parameters if 8×8 parameters are specified. This gives as much transparency as possible to the user, especially when comparing own results with calculations and $\mathbf{k}\cdot\mathbf{p}$ parameters

of published work by others. Additionally, the bulk $\mathbf{k}\cdot\mathbf{p}$ dispersion along the [001], [110] and [111] directions is part of the output, so that the user can check whether the energy dispersion for a particular choice of $\mathbf{k}\cdot\mathbf{p}$ parameters is meaningful, e.g. if the curvature of the energy dispersion has the correct sign (see Fig. 10.1). This is important to check for e.g. alloys, if strain is present or for rescaled $\mathbf{k}\cdot\mathbf{p}$ parameters, or if spurious solutions are present. Finally, we also output the calculated A, B, C parameters that have been used by Dresselhaus, Kip and Kittel [DKK55]. They are of no practical relevance for our software because they only apply along certain symmetry directions in the Brillouin zone where the energies are twofold degenerate, e.g. along the line from L to Γ to X. The DKK (Dresselhaus–Kip–Kittel) equation [DKK55] for the energy dispersion of the holes is valid only for energies small compared to the spin-orbit splitting energy. Nevertheless, they provide insight into the choice of $\mathbf{k}\cdot\mathbf{p}$ parameters. From these parameters, the energy dispersion for the heavy and light holes are obtained

$$E_{\mathrm{hh}}(\mathbf{k}) = Ak^2 + \sqrt{B^2k^4 + C^2\left(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2\right)} \tag{3.135}$$

$$E_{\mathrm{lh}}(\mathbf{k}) = Ak^2 - \sqrt{B^2k^4 + C^2\left(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2\right)}. \tag{3.136}$$

The anisotropy is caused by C. Therefore the constant energy surfaces in reciprocal $\mathbf{k}$ space are 'warped spheres' and show a cubic symmetry. This deviation from the spherical symmetry is a direct consequence of the cubic crystal system. If C is zero, then the energy dispersion is isotropic (spherical approximation). If both, B and C are nonzero, the dispersion is nonparabolic. Both, A and B are negative with diamond being an exception, according to Yu and Cardona [YC99], although there is still uncertainty in the choice of parameters for diamond, see Section 10.1. A, B and C are related to the Luttinger and DKK parameters as follows

$$A = -\gamma_1\frac{\hbar^2}{2m_0} = \frac{L+2M}{3} + \frac{\hbar^2}{2m_0} \tag{3.137}$$

$$B = -2\gamma_2\frac{\hbar^2}{2m_0} = \frac{L-M}{3} \tag{3.138}$$

$$C^2 = 12\left(\gamma_3^2 - \gamma_2^2\right)\left(\frac{\hbar^2}{2m_0}\right)^2 = \frac{1}{3}(N^2 - (L-M)^2). \tag{3.139}$$

Usually, the A, B, C and L, M, N parameters are given in $\frac{\hbar^2}{2m_0}$ units, and the Luttinger parameters in dimensionless units although some authors use different conventions, e.g. atomic units where $\frac{\hbar^2}{2m_0}$ leads to a factor of 1/2 in the equations. Also opposite sign conventions for A, B, C are used, thus very often only $|A|$, $|B|$, $|C|$ are listed, and in fact only the sign of A, which is obvious, is relevant. If A and B are defined as in eq. (3.135) and eq. (3.136), it is clear that A must be negative, but the signs for B and C are not determined from B^2 or C^2, respectively. Therefore, for the inverse relations it is not so obvious to assign the correct sign to γ_2 and γ_3, as the signs for B and C are not well

defined

$$\gamma_1 = -A\frac{2m_0}{\hbar^2} \tag{3.140}$$

$$\gamma_2 = -\frac{B}{2}\frac{2m_0}{\hbar^2} \tag{3.141}$$

$$\gamma_3^2 = \left(\frac{B^2}{4} + \frac{C^2}{12}\right)\left(\frac{2m_0}{\hbar^2}\right)^2 . \tag{3.142}$$

Usually for all cubic group IV, III-V and II-VI materials that we included in the next**nano** database, all three 6×6 $\mathbf{k}\cdot\mathbf{p}$ Luttinger parameters are positive. The only exceptions from this rule, that we are aware of, are the group IV materials C (diamond), Sn and the II-VI materials HgS, HgSe, HgTe. For all these materials apart from diamond all three Luttinger parameters are negative. Apart from diamond, all have an unusual band structure having either a zero band gap or a negative band gap. For diamond several sets of Luttinger parameters exist (see Section 10.1). The signs of the Luttinger parameters vary among these sets. A possible reason for this is, to our believe, that some of the Luttinger parameters have been calculated from A, B and C.

It is often convenient to know estimates of effective masses in particular directions or averaged over all directions. They can be obtained by projection of the angular momentum operators onto a coordinate system that contains the desired direction. Within next**nano**[3], we write out the effective heavy and light hole masses along the [001], [110] and [111] directions. They can be extracted from the Luttinger parameters [HF63, VMRM01]

$$\frac{m_0}{m_{\mathrm{hh}}^{[001]}} = \gamma_1 - 2\gamma_2 = \frac{2m_0}{\hbar^2}(-A+B) \tag{3.143}$$

$$\frac{m_0}{m_{\mathrm{lh}}^{[001]}} = \gamma_1 + 2\gamma_2 = \frac{2m_0}{\hbar^2}(-A-B) \tag{3.144}$$

$$\frac{m_0}{m_{\mathrm{hh}}^{[111]}} = \gamma_1 - 2\gamma_3 \tag{3.145}$$

$$\frac{m_0}{m_{\mathrm{lh}}^{[111]}} = \gamma_1 + 2\gamma_3 \tag{3.146}$$

$$\frac{m_0}{m_{\mathrm{hh}}^{[110]}} = \gamma_1 - \frac{1}{2}\left(\gamma_2 + 3\gamma_3\right) \tag{3.147}$$

$$\frac{m_0}{m_{\mathrm{lh}}^{[110]}} = \gamma_1 + \frac{1}{2}\left(\gamma_2 + 3\gamma_3\right). \tag{3.148}$$

Also the isotropic, averaged heavy, light and split-off hole masses can be derived. For the latter, two approximations are commonly employed [YC99, VMRM01]. The first

one is

$$\frac{m_0}{m_{\mathrm{hh,av}}} = \frac{2m_0}{\hbar^2}\left(-A + B\left(1 + \frac{2C^2}{15B^2}\right)\right) \tag{3.149}$$

$$\frac{m_0}{m_{\mathrm{lh,av}}} = \frac{2m_0}{\hbar^2}\left(-A - B\left(1 + \frac{2C^2}{15B^2}\right)\right) \tag{3.150}$$

$$\frac{m_0}{m_{\mathrm{so,av}}} = \gamma_1 - \frac{E_{\mathrm{P}}\Delta_{\mathrm{so}}}{3E_{\mathrm{gap}}(E_{\mathrm{gap}} + \Delta_{\mathrm{so}})}. \tag{3.151}$$

The second one uses instead of eq. (3.149) and eq. (3.150)

$$\frac{m_0}{m_{\mathrm{hh,av}}} = \frac{2m_0}{\hbar^2}\left(-A + \frac{2}{5}B\left(1 + \frac{3}{2}\sqrt{1 + \frac{4C^2}{9B^2}}\right)\right) \tag{3.152}$$

$$\frac{m_0}{m_{\mathrm{lh,av}}} = \frac{2m_0}{\hbar^2}\left(-A - \frac{2}{5}B\left(1 + \frac{3}{2}\sqrt{1 + \frac{4C^2}{9B^2}}\right)\right). \tag{3.153}$$

Both approximations become identical if $C = 0$, corresponding to negligible warping [YC99].

Finally, we remark that it is sometimes useful to compare the numerical implementation of the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian to analytical or numerical single-band ('effective-mass') results at $\mathbf{k} = 0$, where the energy dispersion is isotropic and parabolic, and described by an effective mass m. By setting $E_{\mathrm{P}} = 0\,\mathrm{eV}$, one decouples the electrons from the holes. Then the 8×8 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian effectively becomes a 6×6 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian for the holes and a single-band Hamiltonian for the electrons, the latter being twofold degenerate due to spin. To be consistent, one then has to use the 6×6 $\mathbf{k}\cdot\mathbf{p}$ parameters for the holes (L, M, N^+, N^- rather than L', M', $N^{+\prime}$, $N^{-\prime}$), and the parabolic single-band effective mass m_{e} of the electron. This is achieved by setting $S = \frac{m_0}{m_{\mathrm{e}}}$. To obtain a dispersion for the holes that is both isotropic and parabolic, it requires us to set $L = M$, $N^+ = 0$ and $N^- = 0$ which implies $N = 0$. This is equivalent to setting $\gamma_2 = \gamma_3 = 0$, and $\kappa = -\frac{1}{3}$, or setting $F = G$ and $H_1 = H_2 = \frac{3}{2}F$. For instance, if we want to achieve a dispersion corresponding to an effective mass of $m_{\mathrm{h}} = 0.5m_0$ for each of the three hole bands, where the split-off band is separated from the degenerate heavy and light hole band energies by the spin-orbit splitting energy Δ_{so}, our $\mathbf{k}\cdot\mathbf{p}$ parameters must be given by

- $\gamma_1 = \frac{m_0}{m_{\mathrm{h}}} = 2$, $\gamma_2 = \gamma_3 = 0$, $\kappa = -\frac{1}{3}$ or
- $L = M = (-\gamma_1 - 1)\frac{\hbar^2}{2m_0} = -3\frac{\hbar^2}{2m_0}$, $N^+ = 0$, $N^- = 0$ or
- $F = G = \frac{L}{3} = -\frac{\hbar^2}{2m_0}$, $H_1 = H_2 = \frac{L}{2} = -\frac{3}{2}\frac{\hbar^2}{2m_0}$.
- (For the Foreman parameters it follows $\sigma = \delta = 1/6$ and $\pi = 1/4$. However, they implicitly assume $H_2 = 0$. This is definitely not the case here because now $H_2 \neq 0$ and its contribution is even larger than the one of F or G.)

These conditions are valid for bulk. They are also valid for heterostructures because we specified four parameters, except for the Foreman parameters. For bulk (without magnetic field), it is sufficient to specify only three parameters, and thus we can also use the Foreman parameters. To obtain an isotropic dispersion for the holes, it requires us to set $N = L - M$, i.e. $N^+ = L - 2M$ and $N^- = M$ (spherical approximation). This is equivalent to setting $\gamma_2 = \gamma_3$, leading to

$$\kappa = -\frac{1}{6}(L - 3M)\frac{2m_0}{\hbar^2} - \frac{1}{3} = -\frac{1}{3}(\gamma_1 - 5\gamma_2 + 2). \tag{3.154}$$

For instance, setting $M = 0$, $N = L$, i.e. $N^+ = L$ and $N^- = 0$ yields an isotropic dispersion. This is equivalent to $\gamma_2 = \gamma_3 = \frac{1}{2}(\gamma_1 + 1) = -\frac{1}{6}L\frac{2m_0}{\hbar^2}$. The spherical approximation is also obtained by replacing γ_2 and γ_3 by

$$\hat{\gamma} = \frac{1}{5}(2\gamma_2 + 3\gamma_3). \tag{3.155}$$

Replacing γ_2 and γ_3 by

$$\tilde{\gamma} = \frac{1}{2}(\gamma_2 + \gamma_3), \tag{3.156}$$

yields the axial approximation, which is characterized by a cylindrical symmetry of the Hamiltonian, i.e. axial symmetry in the (x, y) plane.

If it holds $L = M$ (or $N = 0$), i.e. $N^+ = -M = -L$ and $N^- = M = L$, the hole dispersion is both isotropic and parabolic. This is equivalent to setting $\gamma_2 = \gamma_3 = 0$, leading to

$$\kappa = \frac{1}{3}L\frac{2m_0}{\hbar^2} - \frac{1}{3} = -\frac{1}{3}(\gamma_1 + 2) = -F\frac{2m_0}{\hbar^2} - \frac{1}{3} = 6\sigma - \frac{1}{3}. \tag{3.157}$$

If γ_1 is given, the four sets of parameters are related through $L = (-\gamma_1 - 1)\frac{\hbar^2}{2m_0}$, $F = -\frac{1}{3}L$ and $\sigma = \frac{L}{18}\frac{2m_0}{\hbar^2}$. In this case it holds $G = -2F = \frac{2}{3}L$, $H_1 = -3F = L$ and $H_2 = 0$. The consistent Foreman parameters are $\delta = -2\sigma = -\frac{1}{9}L\frac{2m_0}{\hbar^2}$ and $\pi = -3\sigma = -\frac{1}{6}L\frac{2m_0}{\hbar^2}$, implicitly assuming $H_2 = 0$. For these conditions it further holds $A = -\gamma_1\frac{\hbar^2}{2m_0}$, and $B = C^2 = 0$.

Summary The zinc blende 8×8 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian can be parametrized by the following 10 quantities,

- the band gap energy E_{gap} at the Γ point,
- the spin-orbit split-off energy Δ_{so} (Δ_2, Δ_3, crystal field splitting Δ_1),
- the optical momentum matrix element E_{P} (E_{P1}, E_{P2}),
- the energy of the average of the three valence band edges $E_{\mathrm{v,av}}$,
- the conduction band mass at the Γ point m_{e} ($m_{\mathrm{e},\parallel}$, $m_{\mathrm{e},\perp}$),
- the Luttinger parameters γ_1, γ_2, γ_3, κ (A_1, A_2, A_3, A_4, A_5, A_6, A_7),
- the inversion asymmetry parameter B (B_1, B_2, B_3),

where the 19 wurtzite parameters are given in parenthesis if different from the zinc blende ones. To account for strain effects 4 additional parameters are required for zinc blende, and 7 for wurtzite (see Ref. [Hac02] for details),

- the hydrostatic conduction band deformation potential a_{c}^{Γ} ($a_{\mathrm{c},\parallel}^{\Gamma}$, $a_{\mathrm{c},\perp}^{\Gamma}$),
- the valence band deformation potentials $a_{\mathrm{v,av}}$, b, d (d_1, d_2, d_3, d_4, d_5, d_6).

3.2. Spurious solutions

The numerical discretization of the single-band and $\mathbf{k}\cdot\mathbf{p}$ Schrödinger equations have been discussed in detail in Refs. [Hac02, And04]. Everybody that numerically implements a $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian will sooner or later encounter 'spurious solutions'. These are unphysical wave functions that look strange, e.g. if they oscillate very strongly or if they have spikes at material interfaces. Occasionally, their energies are even lying within the forbidden band gap. These solutions might arise from incorrect operator ordering (i.e. incorrect discretization) or from the used $\mathbf{k}\cdot\mathbf{p}$ parameters. A detailed discussion on spurious solutions can be found in Ref. [And04]. Veprek et al. [VSW07] related the spurious solution problem to the loss of ellipticity of the differential operator. They derived a criteria that must be fulfilled by the $\mathbf{k}\cdot\mathbf{p}$ parameters to establish ellipticity. For all materials used in a calculation, next**nano**^{3} automatically outputs this information so that one can check if this criteria is fulfilled. Foreman [For97] suggested to get rid of spurious solutions by setting $S = 0$ (eq. (3.62)). This requires a rescaling of the E_P parameter

$$E_\mathrm{P} = \left(\frac{m_0}{m_\mathrm{c}^*}\right) \frac{E_\mathrm{gap}\left(E_\mathrm{gap} + \Delta_\mathrm{so}\right)}{E_\mathrm{gap} + \frac{2}{3}\Delta_\mathrm{so}}, \tag{3.158}$$

in order to still get the correct conduction band dispersion. Now E_P is fitted to the electron mass, rather than having S as the fitting parameter. Essentially this implies that remote-band contributions cancel the free-electron term. An alternative is to set $S = 1$ which corresponds to entirely neglecting remote bands. This is the default implementation of the next**nano**++ software. The appropriate equation for E_P then reads

$$E_\mathrm{P} = \left(\frac{m_0}{m_\mathrm{c}^*} - 1\right) \frac{E_\mathrm{gap}\left(E_\mathrm{gap} + \Delta_\mathrm{so}\right)}{E_\mathrm{gap} + \frac{2}{3}\Delta_\mathrm{so}}. \tag{3.159}$$

To be consistent, the modified DKK (eq. (3.101), eq. (3.103), eq. (3.105)) or the modified Luttinger parameters (eq. (3.118), eq. (3.119), eq. (3.120), eq. (3.121)) have to be recalculated using the new value of E_P. Analogous equations for the wurtzite case can be obtained from eq. (3.63) and eq. (3.64).

According to Andlauer [And04], for $(L' + 1)\, S < 0$ no oscillatory spurious solutions exist. He discussed spurious solutions for one-dimensional structures in Ref. [And09]. Here we give the details of a simple two-dimensional example so that anyone interested in spurious solutions can reproduce these results. We have verified that next**nano**^{3}, next**nano**++ and the next**nano**.net software produce the same spurious results for this example. A quadratic InAs quantum wire of dimensions $5\,\mathrm{nm} \times 5\,\mathrm{nm}$ is surrounded by a GaAs barrier of 10 nm thickness. The $\mathbf{k}\cdot\mathbf{p}$ parameters (without rescaling) for InAs are $\gamma_1 = 20.0$, $\gamma_2 = 8.5$, $\gamma_3 = 9.2$, $m_c = 0.026 m_0$, $\Delta_\mathrm{so} = 0.39\,\mathrm{eV}$, $E_\mathrm{gap} = 0.417\,\mathrm{eV}$, $E_\mathrm{P} = 28.8\,\mathrm{eV}$ and the ones for GaAs are $\gamma_1 = 6.98$, $\gamma_2 = 2.06$, $\gamma_3 = 2.93$, $m_c = 0.067 m_0$, $\Delta_\mathrm{so} = 0.341\,\mathrm{eV}$, $E_\mathrm{gap} = 1.519\,\mathrm{eV}$, $E_\mathrm{P} = 21.5\,\mathrm{eV}$. The actual parameters that were used have been rescaled so that $S = 1$ (eq. (3.159)) following the suggestions described above. For the valence band offset between InAs and GaAs we took $E_\mathrm{v,av}^\mathrm{VBO} = 0.044\,\mathrm{eV}$. This is the offset with respect to the average energy of all three hole band edges. The

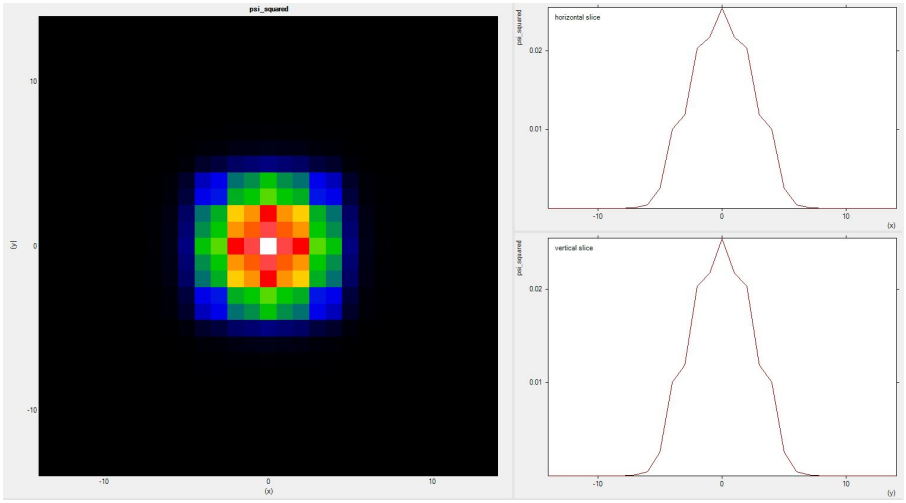


Figure 3.2.: Probability density $\psi_1^2(x, y)$ of the ground state of a square InAs nanowire. Also shown are horizontal and vertical slices through the center.

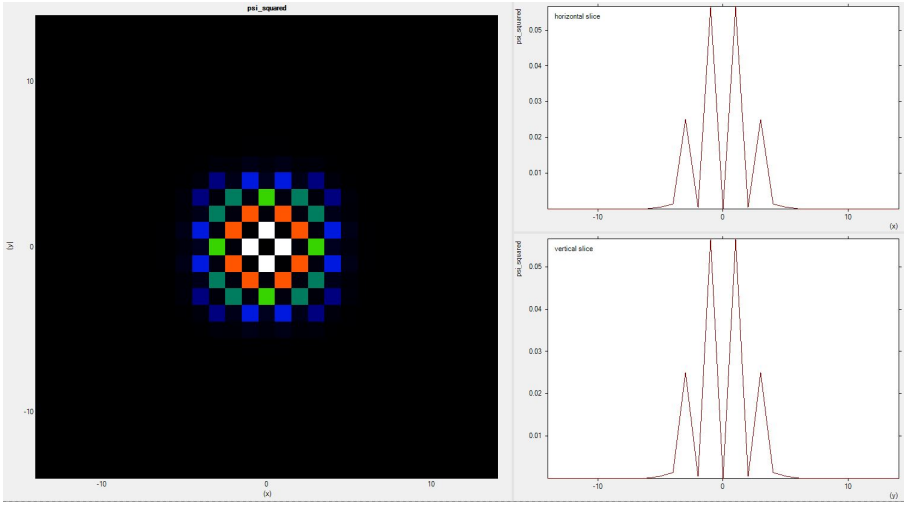


Figure 3.3.: Probability density $\psi_2^2(x, y)$ of the first excited state of a square InAs nanowire. This spurious solution shows a very pronounced oscillatory behavior. The black squares indicate zero probability density.

dimension N of the 8×8 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian matrix is $N = 6728$ corresponding to a total of 29×29 grid points with a grid spacing of 1.0 nm in each direction. Figure 3.2 shows the probability density $\psi_1^2(x, y)$ of the ground state of this square InAs nanowire. While the 2D plot looks kind of okay, the horizontal and vertical slices through the center reveal some kinks in the density which should not be there. The energies of the ground state at $E_1 = 0.138\,\text{eV}$ (twofold degenerate) and of the excited state at $E_2 = 0.220\,\text{eV}$ (fourfold degenerate) are within a reasonable range. Figure 3.3 shows the probability density $\psi_2^2(x, y)$ of the first excited state of this square InAs nanowire. This spurious solution is oscillating very strongly which can easily be seen. The black squares in the 2D plot indicate zero probability density.

It is very challenging to develop a self-consistent $\mathbf{k} \cdot \mathbf{p}$ Schrödinger–Poisson–current solver that will detect automatically spurious solutions if the eigenenergies are within a reasonable range (i.e. not within the band gap) and if relevant states that contribute to the density (e.g. the ground state in our example) show a 'reasonable' probability density, especially if the band edge profile is complicated due to alloy profiles, doping profiles, strain or piezoelectric fields. Rather than having a convenient 'black-box tool', the user must always check if the wave functions are reasonable. There are also situations possible where the probability density looks perfectly okay but the wave function itself is spurious [TES]. In this case one is probably not able to recognize the spurious solution by looking at the density, band edge profile or energy levels. However, quantities that depend on the correctness of the wave functions, like matrix elements or calculated optical absorption spectra will likely show strange results. It is challenging to detect this. Recent progress on the topic of spurious solutions has been made by Eißfeller and Vogl. They developed a spurious-solution-free real-space multi-band envelope function approach that they termed 'symmetry adapted finite element method' [EV11].

3.3. Energy levels in unipolar devices based on intersubband transitions

Figure 3.4 shows a comparison between the widely used single-band effective-mass model and the more sophisticated 8-band $\mathbf{k} \cdot \mathbf{p}$ model for the electron eigenstates of a single (a), double (b) and triple quantum well (c). The geometry, material parameters and doping profiles of these structures are based on Ref. [SCF94] with the exception of the $\mathbf{k} \cdot \mathbf{p}$ parameters which are taken from Ref. [VMRM01] including bowing terms. The QW material consists of $In_{0.53}Ga_{0.47}As$ and the barrier material of $Al_{0.48}In_{0.52}As$. Their alloy contents have been chosen so that both materials are lattice-matched to the underlying InP substrate. Thus we can safely neglect strain in this example. The structures are weakly doped, so band bending is hardly recognizable. The band profiles have been obtained by a self-consistent single-band Schrödinger–Poisson calculation and were used as the potential profiles for the subsequent $\mathbf{k} \cdot \mathbf{p}$ calculation. One can clearly see that for the ground state, the single-band approximation is acceptable. However, the single-band model overestimates intersubband transition energies because for the states further above the conduction band edge, the assumption of parabolic dispersion becomes inac-

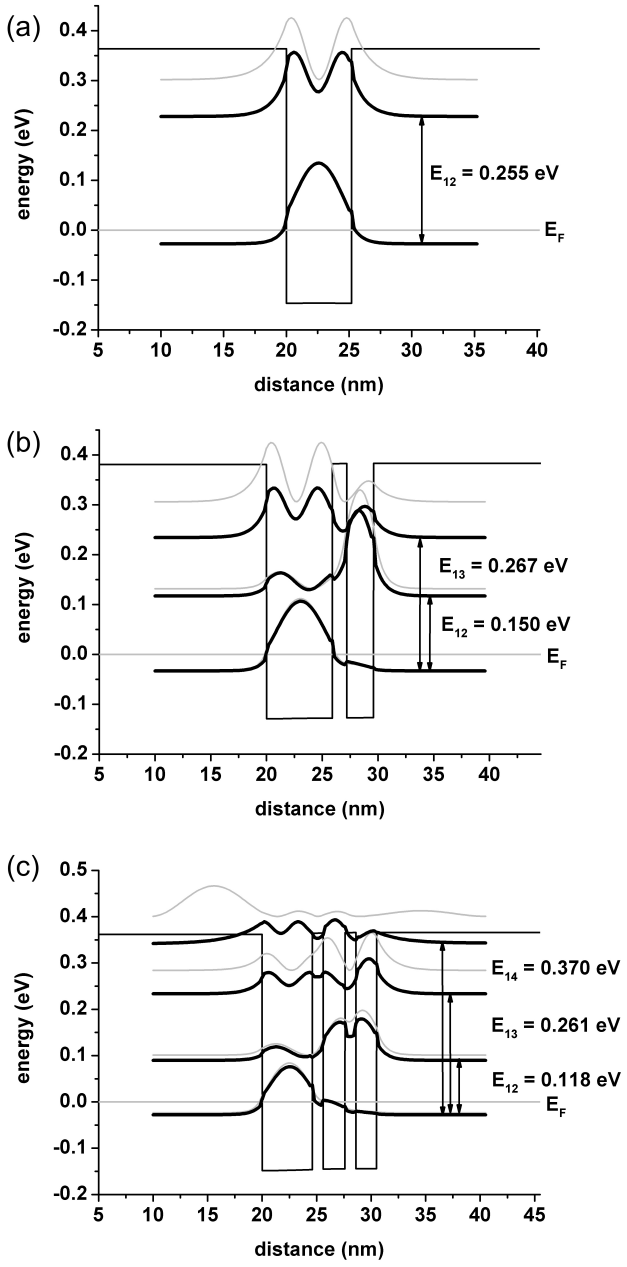


Figure 3.4.: Lowest electron eigenstates of a single (a), double (b) and triple quantum well (c) calculated with an 8-band $\mathbf{k} \cdot \mathbf{p}$ model (thick black lines). The thin gray lines show the eigenstates obtained with a single-band effective mass model. Only for the ground state, these two models are in good agreement but differ significantly for the higher-lying states. In all three cases, the ground state is lying below the Fermi level $E_\mathrm{F} = 0\,\mathrm{eV}$ and thus dominates the charge density. The calculated transition energies are indicated.

curate. Consequently, for unipolar devices that are based on intersubband transitions like quantum cascade lasers (QCL) or quantum well infrared photodetectors (QWIP), a model that takes into account the nonparabolicity of the effective mass is mandatory to get correct energies for the higher-lying states. Our calculated energies for intersubband transitions are in good agreement with the experimental measurements and theoretical calculations of Ref. [SCF94]. For the triple quantum well, our values for the intersubband transition energies of $E_{12} = 118\,\text{meV}, E_{13} = 261\,\text{meV}$ and $E_{14} = 370\,\text{meV}$ compare well with both, the calculated values of Sirtori et al. (116 meV, 257 meV and 368 meV) and their measured values (compare with absorption curve in Fig. 5 of Ref. [SCF94]). For the double quantum well our results are $E_{12} = 150\,\text{meV}$ and $E_{13} = 267\,\text{meV}$ (Sirtori: 150 meV and 271 meV) and for the single quantum well the transition energy is $E_{12} = 255\,\text{meV}$ (Sirtori: 258 meV). For strained structures where the nonparabolicity and anisotropy of the effective masses can change dramatically, the deviations between the single-band and the $\mathbf{k} \cdot \mathbf{p}$ model are even more pronounced.

A common procedure for estimating the probability of intersubband transitions in QCLs or QWIPs between initial state i and final state f, is to evaluate the intersubband dipole moments M_{fi}

$$|M_{fi}| = \left| \int \psi_f^*(x) p_x \psi_i(x) \mathrm{d}x \right| = \left| -i\hbar \int \psi_f^*(x) \frac{\mathrm{d}}{\mathrm{d}x} \psi_i(x) \mathrm{d}x \right|, \tag{3.160}$$

where the structure is assumed to be grown along the x direction, and ψ_f and ψ_i are the envelope wave functions involved in the transitions of interest. In order to model and optimize QCL structures, the four main outputs of a single-band Schrödinger solver, typically used in the QCL community, are wave functions, energy levels, intersubband transition matrix elements, and LO phonon scattering rates (lifetimes) which are implemented in next**nano**^{3}. The latter is based on Ref. [FB89] and is described in detail in Ref. [Sca02]. Only recently, more sophisticated approaches based on the nonequilibrium Green's function method (NEGF) have been developed [Kub09], that gain much further insight into relevant QCL operation principles, like e.g. the concrete influence of LO phonon scattering and population inversion. The NEGF approach is sketched in Ref. [BKV08]. An example of a quantum cascade structure is shown in Fig. 3.5. The SiGe–Si design is based on Ref. [DDG$^+$00]. The valence band edges of the heavy hole and light hole at an electric field of 50 kV/cm are indicated by the black and red thin solid lines. In Si, they are degenerate whereas in SiGe they are nondegenerate due to strain with respect to a Si substrate. We solved the 6×6 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian including strain. The probability densities of the calculated eigenstates are shifted by their energies. Most of the states are shown in thin gray lines. Some of the states of the injector and collector regions are shown in color, highlighting the periodicity of the structure. The calculated electronic transition energy between the upper state (labeled with HH1, thick black solid line) and the lower state (HH2, thick red solid line) of the vertical transition in the active SiGe QW of width 4 nm is 124.5 meV, which matches surprisingly well to the experimental result of 125 meV [DDG$^+$00]. We emphasize that we used the default parameters of the next**nano**^{3} database, so there was no fitting of material parameters to match the experiment. This is indeed remarkable, because a lot

of material parameters were involved, namely lattice constants and associated temperature dependence constants ($T = 50\,\mathrm{K}$ in this example), DKK parameters, spin-orbit splitting energies, elastic constants, and valence band deformation potentials of Si and Ge, as well as the valence band offset between Si and Ge. Additionally, in the quantum wells, the respective linearly interpolated material parameters for the $Si_{0.68}Ge_{0.32}$ and $Si_{0.79}Ge_{0.21}$ alloys were involved.

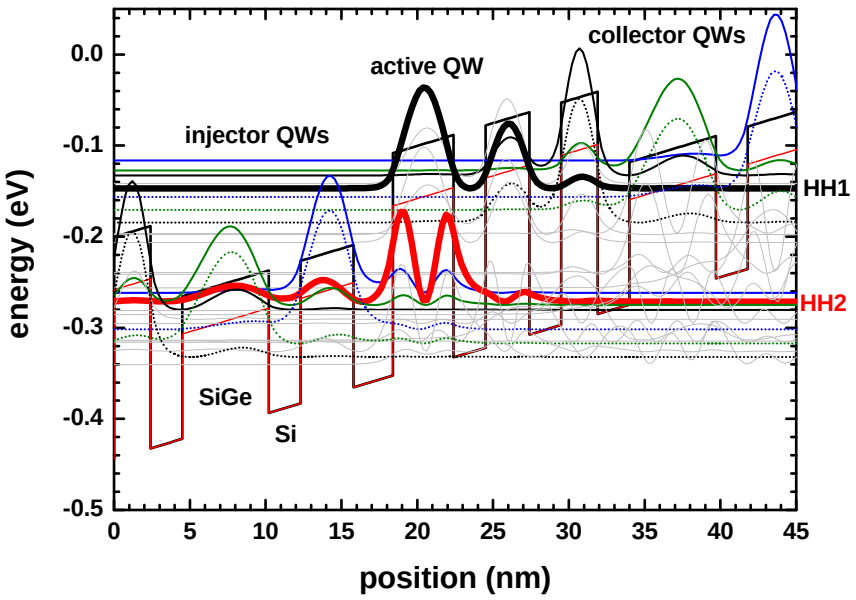


Figure 3.5.: Probability densities, shifted by their eigenenergies, and valence band edges (black and red thin solid lines) of a strained SiGe–Si quantum cascade structure. The vertical transition in the active SiGe QW takes place between the HH1 state (upper state, thick black solid line) and the HH2 state (lower state, thick red solid line).

3.4. Spin-orbit coupling in silicon quantum dots

It is difficult to compare numerical results of a three-dimensional $\mathbf{k} \cdot \mathbf{p}$ calculation to analytical results due to its complexity. Thus one needs simple model systems where one can compare the results of different numerical implementations against each other. In this example we demonstrate the effect of spin-orbit coupling on the degeneracy of the eigenstates of a spherical silicon quantum dot (QD) with a diameter of 5 nm. We solve the 6×6 $\mathbf{k} \cdot \mathbf{p}$ Schrödinger equation for the hole eigenstates where we set the spin-orbit coupling energy either to $\Delta_{\mathrm{so}} = 44\,\mathrm{meV}$ or to zero. For simplicity we assume

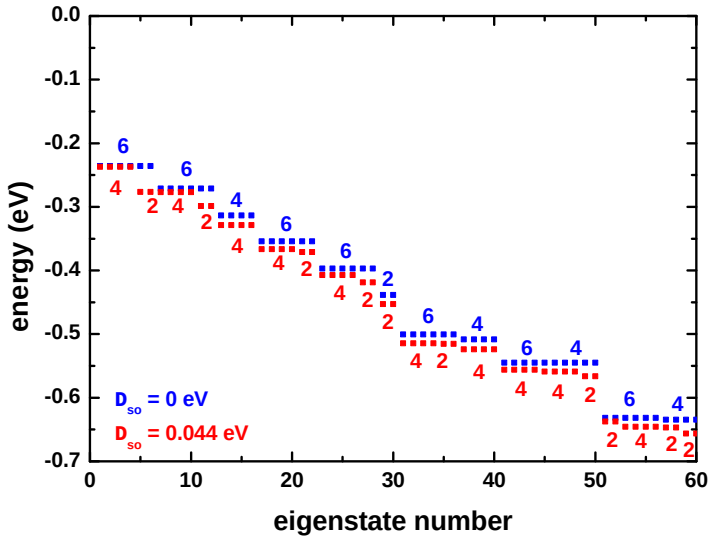


Figure 3.6.: Hole energy levels of a spherical silicon quantum dot of diameter 5 nm with (blue squares) and without (red squares) spin-orbit coupling calculated with the $\mathbf{k} \cdot \mathbf{p}$ method. The degeneracies of the levels are indicated by numbers.

infinite barriers at the QD boundaries which is a reasonable approximation as silicon nanocrystals are typically surrounded by a SiO_2 shell with a large valence band offset around 5 eV. The valence band edge energy inside the QD has been set to be 0 eV. The DKK parameters that have been used are

$$L = -6.8\frac{\hbar^2}{2m_0},\ M = -4.43\frac{\hbar^2}{2m_0},\ N = -8.61\frac{\hbar^2}{2m_0}, \tag{3.161}$$

that correspond to the Luttinger parameters $\gamma_1 = 4.22$, $\gamma_2 = 0.39$ and $\gamma_3 = 1.44$. They represent the anisotropy of the hole dispersion in silicon which is rather strong.

Figure 3.6 shows the energy spectrum of the spherical Si QD with (red squares) and without (blue squares) spin-orbit splitting. Without splitting there is a sixfold degeneracy of the ground state. Spin-orbit splitting reduces this degeneracy to fourfold. In general, each state is twofold degenerate due to spin. Additional geometric degeneracies arise due to the spherical symmetry of the QD. According to Burdov [Bur02] one can calculate the ground state energy for this particular system from the L and M parameters with a high degree of accuracy using

$$E_1 = -\frac{\hbar^2}{2m_\mathrm{h}}\frac{\pi^2}{R^2} \tag{3.162}$$

$$m_\mathrm{h} = \frac{3m_0}{(\frac{2m_0}{\hbar^2}L + 1) + 2(\frac{2m_0}{\hbar^2}M + 1)} = \frac{3m_0}{\frac{2m_0}{\hbar^2}(L + 2M) + 3}, \tag{3.163}$$

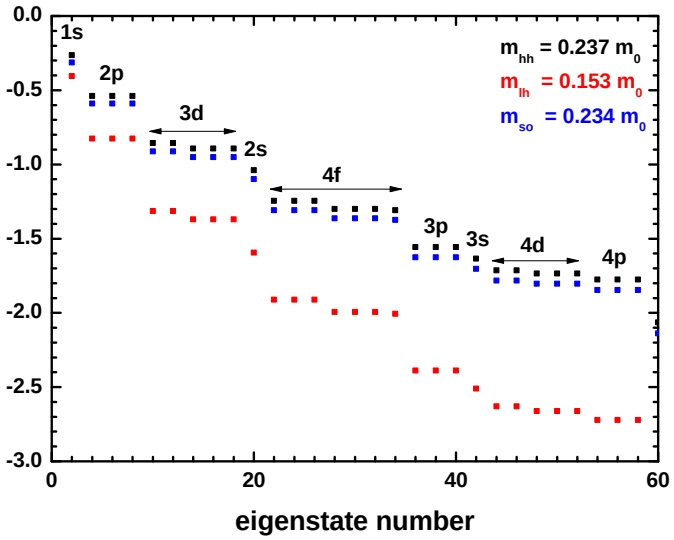


Figure 3.7.: Hole energy levels of a spherical silicon quantum dot ('artificial atom') of diameter 5 nm for heavy (black squares), light (red squares) and split-off holes (blue squares) calculated with the single-band Schrödinger equation. The degeneracies of the levels are indicated by the standard atomic orbital notation known from the electron configuration of atoms.

where R is the radius of the sphere and $m_\mathrm{h} = -0.237 m_0$ is the isotropic hole mass for our choice of parameters with the minus sign for the hole mass indicating that the dispersion in the bulk material is bent downwards. The calculated value for the ground state energy $E_1 = -0.254\,\mathrm{eV}$ is close to our numerical value of $-0.237\,\mathrm{eV}$ using the 6×6 $\mathbf{k} \cdot \mathbf{p}$ method. Burdov writes down eq. (3.163) without the free electron term '+1' which is overall consistent within their article but their parameters [Bur02] are unfortunately incorrect. The parameters given in eq. (3.161) which are derived from the Luttinger parameters of Lawaetz [Law71] are the correct ones whereas the parameters termed L and M in Burdov's article are actually the L^LK and M^LK parameters (eq. (3.33), eq. (3.34)). Thus they have to be corrected by '+1', i.e. $L^\mathrm{LK} = -5.8\frac{\hbar^2}{2m_0}$ and $M^\mathrm{LK} = -4.43\frac{\hbar^2}{2m_0}$ would have been the correct values to be consistent within their article (see also the discussion on the different and thus confusing definitions of the L and M parameters in Section 3.1). Consequently, their calculated hole mass of $m_\mathrm{h} = -0.19 m_0$ has to be corrected by our value of $m_\mathrm{h} = -0.237 m_0$.

For comparison, the results of the single-band calculations with isotropic heavy, light and split-off hole masses are shown in Fig. 3.7 for the same quantum dot. Here, each state is twofold degenerate due to spin but only one of these two energy levels is shown. The numbering of the horizontal axis, however, is taking spin into account. Figure 3.6 shows the lowest 60 eigenvalues. This numbering corresponds roughly to the first 20

eigenvalues for each hole species in Fig. 3.7. The degeneracies of the levels are indicated by the standard atomic orbital notation known from the electron configuration of atoms ($1s$, $2p$, $3d$, $2s$, $4f$, ...). From that labeling it becomes clear why quantum dots are termed 'artificial atoms'. In contrast to real atoms, their energy levels can be tuned to match technologically relevant energy regimes, while showing qualitatively similar energy spectra as atoms, at least for spherical dots. Self-organized quantum dots, however, are mainly showing an energy spectrum similar to a two-dimensional harmonic oscillator as their confinement potential in the plane perpendicular to the growth direction can often be approximated by a parabolic confinement (see also the discussion in Subsection 2.3.1). The fivefold degeneracy of the d levels and the sevenfold degeneracy of the f level is not reproduced well because the ideal shape of the spherical QD is approximated by a QD discretized on a rectangular grid having cubic symmetry. Obviously, the numerical single-band results are very poor in comparison to the $\mathbf{k} \cdot \mathbf{p}$ results with the exception of the ground state energy $E_1 = -0.265\,\mathrm{eV}$.

We have verified that both the next**nano**^{3} and the next**nano**++ software lead to the same $\mathbf{k} \cdot \mathbf{p}$ eigenvalue spectrum (not shown). For both, we used a cuboidal shaped quantum region although next**nano**^{3} is capable of using arbitrarily shaped quantum regions, e.g. a spherical quantum region which is numerically more efficient as less quantum grid points are needed. If m grid points can be excluded from the quantum region due to an optimal choice of quantum region shape, the dimension N of the 6×6 $\mathbf{k} \cdot \mathbf{p}$ matrix reduces to $N - 6m$.

For this particular spherical geometry, the eigenvalues are highly degenerate, not only due to spin but also due to geometry. This sometimes causes problems for certain eigenvalue solvers as they might miss some of these degenerate eigenvalues. For instance, our implementation of the Arnoldi method that uses Chebyshev polynomials as preconditioner [TZA$^+$06] missed some degenerate eigenvalues. For this reason it is of great advantage if any numerical software has redundancy in terms of several eigensolvers, where one can choose from, in order to check results for consistency and accuracy, as well as performance. The ARPACK eigenvalue solver [LSY98] took around 6 minutes for 60 eigenvectors where the dimension of the matrix was $N = 55566$ corresponding to $21 \times 21 \times 21 = 9261$ grid points with a grid resolution of 0.25 nm in each direction.

3.5. Type-III broken-gap band alignment – HgTe–CdTe quantum well

HgTe is an interesting material for studies of the intrinsic spin Hall effect [BRN$^+$10] and the quantum spin hall effect [BHZ06], or spin splitting effects in general due to its large Rashba-type spin-orbit splitting. HgTe is a zero-gap semiconductor that can be embedded between CdTe layers to form a HgTe–CdTe quantum well (QW) heterostructure which shows an interesting type-III band alignment where the valence band edge in the HgTe QW lies above its conduction band edge. Due to this band alignment it is not possible to apply a single-band Hamiltonian. Thus a $\mathbf{k} \cdot \mathbf{p}$ or tight-binding approach is required. Large HgTe quantum wells have an inverted band structure where the highest

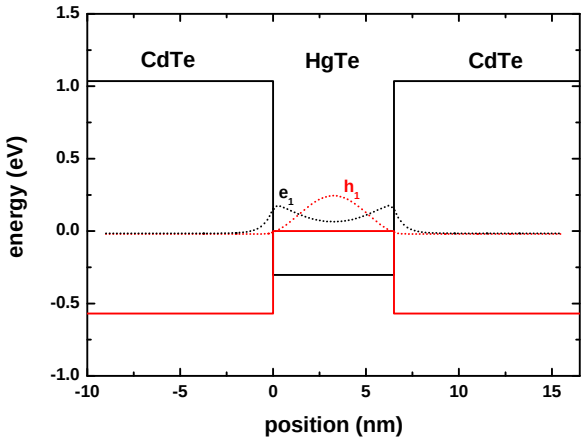


Figure 3.8.: Probability density of the lowest electron (e_1) and highest hole (h_1) eigenstates of a 6.5 nm HgTe quantum well calculated with the $\mathbf{k} \cdot \mathbf{p}$ method. In the $\mathbf{k} \cdot \mathbf{p}$ method, the eigenstates correspond to envelope functions. The conduction (black solid line) and valence band edges (red solid line) form a type-III band alignment.

hole state (h_1) lies above the lowest electron state (e_1). For smaller quantum well widths, the quantum confinement increases and below a critical well width, the band structure becomes normal again with the electron state above the hole state. Figure 3.8 shows the square of the calculated $\mathbf{k} \cdot \mathbf{p}$ wave functions of e_1 and h_1 at the crossover well width at 6.5 nm. Increasing the well width shifts the e_1 state below the h_1 state. This is shown in Fig. 3.9 where the probability density of the relevant states have been calculated with the empirical tight-binding method for a 7.8 nm HgTe quantum well. One can nicely see that in the tight-binding method the envelope of the probability density corresponds to $\mathbf{k} \cdot \mathbf{p}$ envelope functions. For the $sp^3d^5s^*$ tight-binding [JSBB98] calculations, we used a valence band offset of 0.4 eV. For the $\mathbf{k} \cdot \mathbf{p}$ calculations, we took exactly the same material parameters as in Ref. [NPJJ$^+$05], including their valence band offset of 0.570 eV. In both cases, we neglected strain effects for simplicity.

Figure 3.10 shows the energies of the electron and hole states in a HgTe–CdTe quantum well as a function of HgTe QW width calculated with the 8×8 $\mathbf{k} \cdot \mathbf{p}$ method. The crossover of normal to inverted band structure occurs around 6.5 nm and corresponds to the situation in Fig. 3.8. The dashed lines indicate the energetic positions of the conduction and valence band edges of the HgTe QW. Our results for the crossover width are in good agreement to the calculations of Novik et al. [NPJJ$^+$05], and also close to tight-binding calculations [CST].

I have implemented Peter Vogl's TIGHTEN superlattice code into the next**nano**^{3} software package, so that it is now more convenient to perform systematic comparisons between the $\mathbf{k} \cdot \mathbf{p}$ and the tight-binding method for quantum wells.

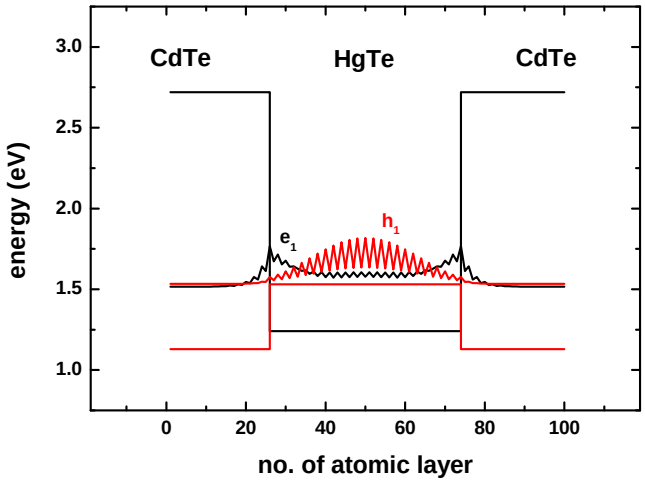


Figure 3.9.: Probability density of the lowest electron (e_1) and highest hole (h_1) eigenstates of a 7.8 nm HgTe quantum well calculated with the empirical tight-binding method.

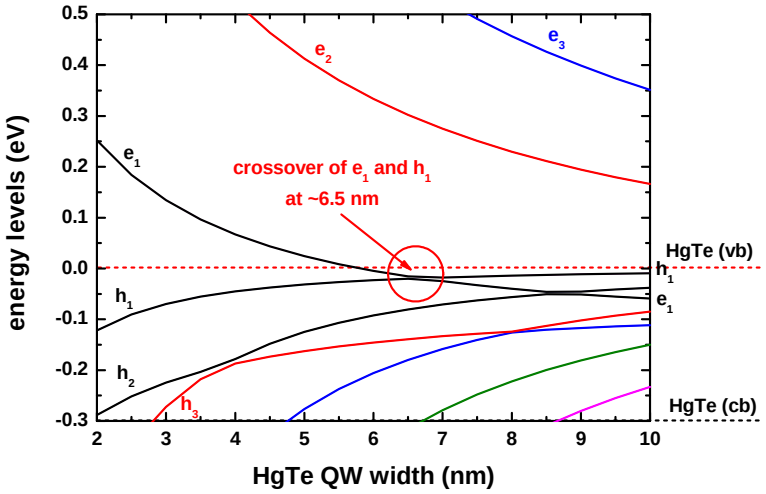


Figure 3.10.: Calculated energies of the electron and hole states in a HgTe–CdTe quantum well as a function of HgTe QW width (8×8 $\mathbf{k} \cdot \mathbf{p}$). The crossover of normal to inverted band structure occurs around 6.5 nm and corresponds to the situation in Fig. 3.8. The dashed lines indicate the conduction and valence band edges of the HgTe QW.

3.6. Type-II broken-gap band alignment – InAs–GaSb superlattice

Type-II broken gap heterostructures have the unusual property that in one material the lowest conduction band edge energy lies below the highest valence band edge energy of the material next to it. Consequently, they cannot be modeled by a single-band model, similar as the type-III structures of the previous section. There will be a strong coupling of the electron and hole wave functions, in some cases even a very strong wave function hybridization, making it eventually impossible to distinguish between electron and hole states if the lowest 'conduction' band state lies below the highest 'valence' band state [GYFE95, ZYC01, AV09]. In this section, we calculate the energy dispersion of an InAs–GaSb superlattice along the in-plane directions $\mathbf{k}_{\|}$ and along the superlattice direction $\mathbf{k}_{\mathrm{SL}}$ by two different methods, the 8×8 $\mathbf{k}\cdot\mathbf{p}$ and the $sp^3d^5s^*$ [JSBB98] tight-binding method. Such structures are relevant for infrared detectors or to study semimetal–semiconductor phase transitions. Remarkably, for certain layer widths the in-plane subband dispersion becomes linear, indicating vanishing effective masses [AV09]. In such a situation, the band structure of the InAs–GaSb superlattice is similar to the band structure around the Dirac point in graphene (see Section 9.1), where a two-dimensional gas of massless Dirac fermions is formed. It is interesting to notice that such properties can also be achieved by growing InAs–GaSb superlattices. However, such topics are not part of our study here.

First, we examine the bulk band structure of InAs along the [110] and [100] directions in $\mathbf{k}$ space at zero temperature. We compare our 8×8 $\mathbf{k}\cdot\mathbf{p}$ model to the single-band effective mass dispersions and to the more sophisticated $sp^3d^5s^*$ tight-binding model which allows one to calculate the bulk band structure in the whole Brillouin zone accurately also for higher-lying conduction bands, e.g. the ones with minima at the X points in the Brillouin zone. The latter is also possible for a $\mathbf{k}\cdot\mathbf{p}$ model where more than eight bands are included [RAF04]. Figure 3.11 shows the results. The parabolic model (dotted lines) is reasonable only for small $\mathbf{k}$ vectors around the Γ point. We used effective masses of $m_{\mathrm{e}}=0.023m_0$, $m_{\mathrm{hh}}=0.41m_0$, $m_{\mathrm{lh}}=0.026m_0$ and $m_{\mathrm{hh}}=0.14m_0$ for the electron, heavy hole, light hole and split-off hole, respectively. The energies for the band gap $E_{\mathrm{gap}}=0.417\,\mathrm{eV}$ and for the split-off energy $\Delta_{\mathrm{so}}=0.38\,\mathrm{eV}$, which are obtained from experiment, are the same in all models and are indicated by the arrows. For the parabolic and the $\mathbf{k}\cdot\mathbf{p}$ model, these values are input parameters while the tight-binding parameters are fitted to yield these values. InAs is a material where the band gap and the split-off energy are of similar energy. This is typically not the case for the group IV or most other III-V materials. Along the [100] direction the energies are twofold degenerate in all models. Due to symmetry arguments, also along the [111] directions the energies are twofold spin degenerate (not shown). However, this spin degeneracy is lifted along the [110] direction for both the $\mathbf{k}\cdot\mathbf{p}$ (black solid lines) and the tight-binding model (red dashed lines). For both, intersubband transitions in InAs quantum wells and density calculations, where the carriers are located mainly around the Gamma point, only small $\mathbf{k}$ values are relevant. For small $\mathbf{k}$ values, the $\mathbf{k}\cdot\mathbf{p}$ energy dispersion is reasonably close

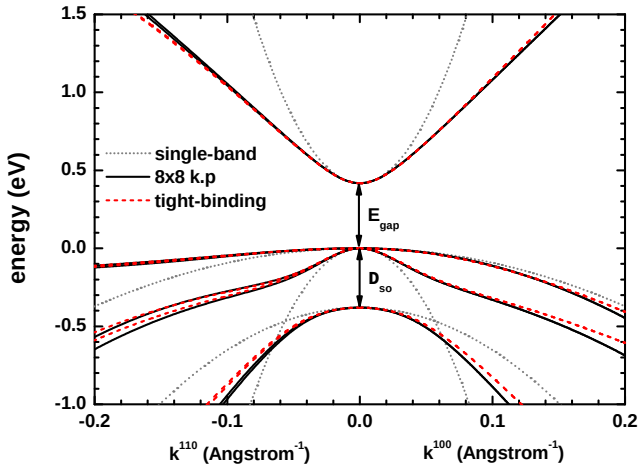


Figure 3.11.: Energy dispersion of bulk InAs along the [110] and [100] directions in $\mathbf{k}$ space calculated with the parabolic model (dotted lines), the 8×8 $\mathbf{k}\cdot\mathbf{p}$ model (black solid lines) and with the $sp^3d^5s^*$ tight-binding parameterization (red dashed lines) at $T = 0\,\mathrm{K}$. Along the [100] direction the energies are twofold degenerate while this spin degeneracy is lifted along the [110] direction for both the $\mathbf{k}\cdot\mathbf{p}$ and the tight-binding model.

to the tight-binding dispersion for both directions, also in terms of spin-splitting. In this calculation, the $\mathbf{k}\cdot\mathbf{p}$ inversion asymmetry parameter B (eq. (3.67)) was taken to be nonzero ($B = 3.60\frac{\hbar^2}{2m_0}$ [Car03]). If it were zero, which is the typical assumption in practically all $\mathbf{k}\cdot\mathbf{p}$ calculations, a spin-splitting along the [110] direction would not be obtained. We conclude that the $\mathbf{k}\cdot\mathbf{p}$ model is sufficiently accurate as a description of the realistic band structure. Of course, the tight-binding approach is also a model but it has been fitted to energies and masses through the entire Brillouin zone, so one can assume that it is a very reasonable model.

We now examine the bulk band structure of biaxially, tensilely strained InAs with respect to a GaSb substrate. The biaxial strain $\epsilon_\parallel = 0.0062$ (eq. (1.10)) is with respect to the (x, y) plane (which we call in-plane) and the strain $\epsilon_\perp = -0.0067$ (eq. (C.6)) is with respect to the z direction (which we call out-of-plane direction). Figure 3.12 shows the energy dispersion of biaxially, tensilely strained InAs along the [110] and [100] directions (in-plane directions) in $\mathbf{k}$ space calculated with the parabolic model (dotted lines), the 8×8 $\mathbf{k}\cdot\mathbf{p}$ model (black solid lines) and with the $sp^3d^5s^*$ tight-binding parameterization (red dashed lines) at $T = 0\,\mathrm{K}$. Along the [100] direction the energies are still twofold degenerate for the $\mathbf{k}\cdot\mathbf{p}$ model, while this spin degeneracy is now lifted for the tight-binding model. The single-band results have the same masses as for unstrained InAs but the band edges are shifted in the same way as for the $\mathbf{k}\cdot\mathbf{p}$ model. The energies of these band edges can easily be obtained by diagonalizing the bulk $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian,

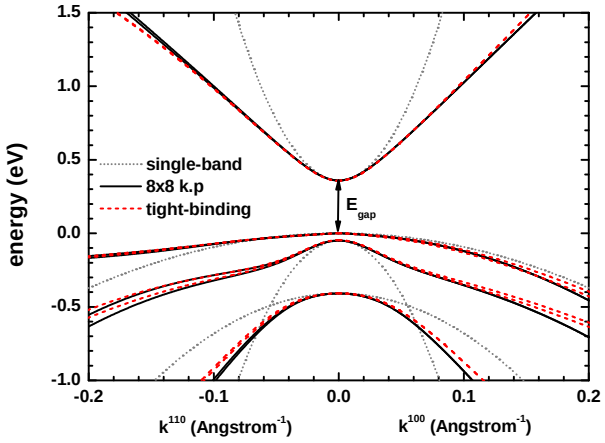


Figure 3.12.: Energy dispersion of biaxially, tensilely strained InAs along the [110] and [100] directions (in-plane directions) calculated with the parabolic model (dotted lines), the 8×8 $\mathbf{k} \cdot \mathbf{p}$ model (black solid lines) and with the $sp^3d^5s^*$ tight-binding parameterization (red dashed lines) at $T = 0\,\mathrm{K}$. Along the [100] direction the energies are still twofold degenerate for the $\mathbf{k} \cdot \mathbf{p}$ model while this spin degeneracy is lifted for the tight-binding model.

that includes the deformation potentials and strain [Hac02], at $\mathbf{k} = 0$. Alternatively, for growth along any direction, analytical equations can be used instead to obtain the shifted and split band edges due to strain for the single-band model [VdW89, PGBD+11]. The common zero point of energy for all three models has been set to the highest hole energy level. The band gap has decreased to $E_{\mathrm{gap}} = 0.359\,\mathrm{eV}$ because the unit cell has increased, corresponding to a positive hydrostatic strain of $\epsilon_{\mathrm{hydro}} = 0.0057$. (If the unit cell increases due to increasing temperature, the band gap also gets smaller.) The degeneracy of the heavy and light hole band edges at the Γ point is now lifted and they are separated by $0.049\,\mathrm{eV}$.

Figure 3.13 shows the same as Fig. 3.12 but along the [001] (out-of-plane) and [100] (in-plane) directions. Along both directions the energies are twofold degenerate for the $\mathbf{k} \cdot \mathbf{p}$ model, while this spin degeneracy is lifted for the tight-binding model. Along the [001] (out-of-plane) direction the dispersion is now very different. The heavy and light hole dispersions cross along the out-of-plane direction. Now the highest hole band is 'light' along the out-of-plane direction, and 'heavy' along the in-plane directions. The first excited hole state is however 'heavy' along the out-of-plane direction, and 'light' along the in-plane directions. Both, the $\mathbf{k} \cdot \mathbf{p}$ and the tight-binding model show very similar results although the dispersion is now rather complicated. This gives further confidence into our $\mathbf{k} \cdot \mathbf{p}$ model.

I have implemented Peter Vogl's TIGHTEN bulk code into the next**nano**^{3} software package. Therefore, it is now possible to calculate with only one input file the $\mathbf{k} \cdot \mathbf{p}$ and

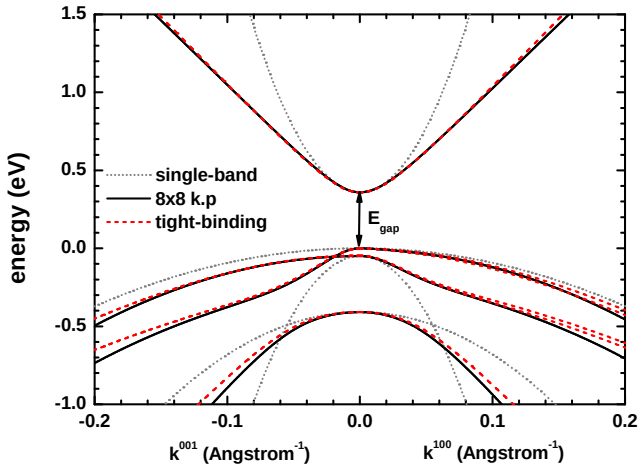


Figure 3.13.: Energy dispersion along the [001] (out-of-plane) and [100] (in-plane) directions for the same situation as in Fig. 3.12. Along the [001] (out-of-plane) direction the dispersion is now very different. Both, the $\mathbf{k} \cdot \mathbf{p}$ (black solid lines) and the tight-binding model (red dashed lines) show very similar results although the dispersion is now rather complicated.

the tight-binding bulk band structure for unstrained or biaxially strained zinc blende materials, and also for ternary alloys, making it very convenient to perform systematic comparisons between the bulk $\mathbf{k} \cdot \mathbf{p}$ and the tight-binding method.

Now we turn to the InAs–GaSb heterostructure. Our structure consists of a 3.7 nm (24 atomic layers) GaSb and a 6.7 nm (44 atomic layers) InAs region. Similar structures have been investigated theoretically with the $\mathbf{k} \cdot \mathbf{p}$ method by e.g. Grein et al. [GYFE95] or Zakharova et al. [ZYC01]. Periodic boundary conditions are used to mimic a superlattice with a periodic length of $L = 10.4\,\mathrm{nm}$ corresponding to $k_{\mathrm{SL,max}} = \pi/L = 0.30\,\mathrm{nm}^{-1}$. Strain has been included assuming that the GaSb layer is unstrained and that the InAs layer is biaxially strained with respect to a GaSb substrate with $\epsilon_{\parallel} = 0.0062$ and $\epsilon_{\perp} = -0.0067$, i.e. InAs is tensilely strained. In order to avoid spurious solutions, the $\mathbf{k} \cdot \mathbf{p}$ material parameters were rescaled according to eq. (3.159) so that $S = 1$ (see Section 3.2). These rescaled parameters were also used in the previous figures for the band structure of bulk InAs. The grid spacing resolution was 0.1 nm in the $\mathbf{k} \cdot \mathbf{p}$ calculation. The tight-binding calculations have been performed with the same method as in the previous section. For both the bulk $\mathbf{k} \cdot \mathbf{p}$ and the tight-binding calculation, the conduction and the three valence band edge energies at $\mathbf{k} = 0$ are identical. This holds for strained InAs (Fig. 3.12, Fig. 3.13) and for GaSb. However, it cannot be avoided that slight deviations in the bulk dispersion occur for nonzero $\mathbf{k}$ vectors due to the different methods employed, in particular if strain is present. Consequently, it is expected that for heterostructure calculations, the $\mathbf{k} \cdot \mathbf{p}$ and tight-binding calculations deviate even more.

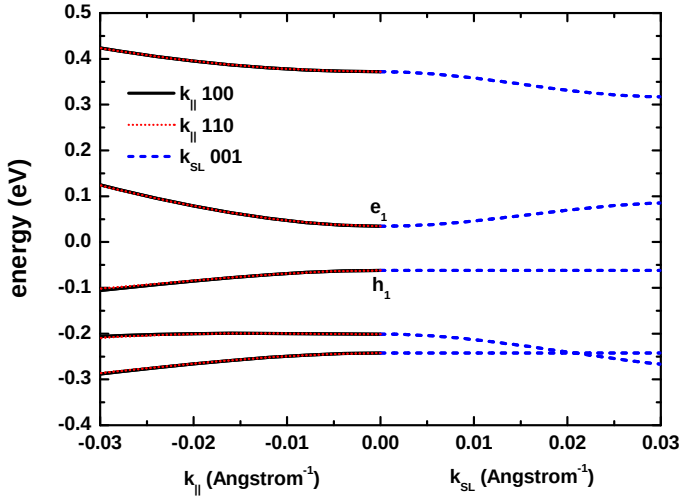


Figure 3.14.: Energy dispersion of a strained InAs–GaSb superlattice along the in-plane directions in $\mathbf{k}$ space, $\mathbf{k}_{||}^{100}$ and $\mathbf{k}_{||}^{110}$, and along the superlattice growth direction $\mathbf{k}_{\mathrm{SL}}$ calculated with the $\mathbf{k} \cdot \mathbf{p}$ method. For small $\mathbf{k}_{||}$ vectors the dispersion is practically isotropic.

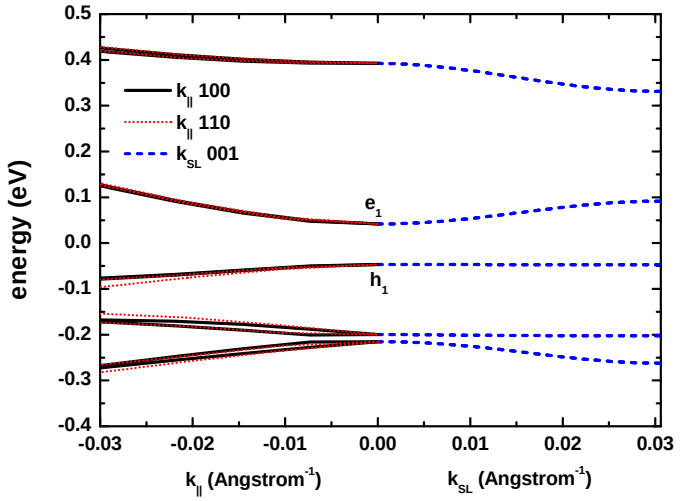


Figure 3.15.: Same as Fig. 3.14 but now showing the results of the tight-binding calculations. The dispersion along $\mathbf{k}_{||}$ is split by strain and shows a slight anisotropy.

Figure 3.14 shows the energy dispersion of the strained InAs–GaSb superlattice along the in-plane directions in $\mathbf{k}$ space, $\mathbf{k}_{\|}^{100}$ and $\mathbf{k}_{\|}^{110}$, and along the superlattice growth direction $\mathbf{k}_{\mathrm{SL}}$ calculated with the 8×8 $\mathbf{k} \cdot \mathbf{p}$ method. For small $\mathbf{k}_{\|}$ vectors the dispersion is practically isotropic. The highest hole state h_1 does not have any curvature along the superlattice direction $\mathbf{k}_{\mathrm{SL}}$. The reason is that the ground state hole wave function is strongly confined in the individual GaSb layers (not shown) and does not couple to neighboring GaSb layers. All other states couple to neighboring layers and thus form minibands (see Section 2.2). The corresponding tight-binding results are shown for comparison in Fig. 3.15. Now the dispersion along $\mathbf{k}_{\|}$ is split by strain and shows a slight anisotropy. The spin-splitting is due to the fact that zinc blende materials do not have inversion symmetry. This is called bulk inversion asymmetry (BIA). Furthermore, the interfaces do not have a common atom, such as the As anion in AlAs–GaAs heterostructures. The latter can have a mirror plane if the number of arsenic layers is odd. Thus for our tight-binding calculations, no mirror plane is present. This is called structural inversion asymmetry (SIA). Consequently, both BIA and SIA were involved. The B parameter that is related to the missing inversion symmetry in zinc blende materials was assumed to be zero in the $\mathbf{k} \cdot \mathbf{p}$ calculations of Fig. 3.14. Structural asymmetry due to the atomic arrangement of the interfaces cannot be taken into account by the $\mathbf{k} \cdot \mathbf{p}$ method. However, spin-splitting due to structural asymmetry introduced by e.g. an electric field or by a heterostructure that does not have an inversion center, is automatically taken into account by the $\mathbf{k} \cdot \mathbf{p}$ method. Within the $\mathbf{k} \cdot \mathbf{p}$ model, our superlattice structure has a mirror plane, i.e. our structure is symmetric, therefore spin-splitting cannot be obtained here because SIA is absent, and BIA was ignored. Both the tight-binding and the $\mathbf{k} \cdot \mathbf{p}$ calculations show qualitatively very similar results, apart from the crossing of the second and third holes states along the superlattice direction. Also the energy levels look very similar. There are rare cases in the literature where the B parameter was actually set to a nonzero value. Typically it is said that it can be neglected or that the parameter is not known. Cartoixà [Car03] lists values for a few III-V materials. In Fig. 3.16 we used his values of $B = 13.1\frac{\hbar^2}{2m_0}$ (GaSb) and $B = 3.60\frac{\hbar^2}{2m_0}$ (InAs). The isotropic dispersion for small $\mathbf{k}_{\|}$ vectors, that we had for $B = 0$, is now slightly lifted. Also the twofold spin-degeneracy along the $\mathbf{k}_{\|}$ directions is no longer present. The order of magnitude of the spin-splitting is comparable to the tight-binding results. Finally, we show the energy dispersion of a strained InAs–GaSb superlattice along the in-plane directions $\mathbf{k}_{\|}^{100}$ and $\mathbf{k}_{\|}^{110}$ calculated with the $\mathbf{k} \cdot \mathbf{p}$ method, where the inversion asymmetry parameter B is zero (black solid lines), and where it is nonzero (red dotted lines) to highlight the differences. In the latter case the twofold spin degeneracy of the energy levels for nonzero $\mathbf{k}_{\|}$ is lifted due to the bulk inversion asymmetry parameter B as discussed above. Therefore the results are closer to the tight-binding calculations. We conclude that it is important to use a nonzero value for the inversion asymmetry parameter B in order to get more realistic results. Further examples on superlattice dispersions using the single-band model were discussed in Section 2.2.

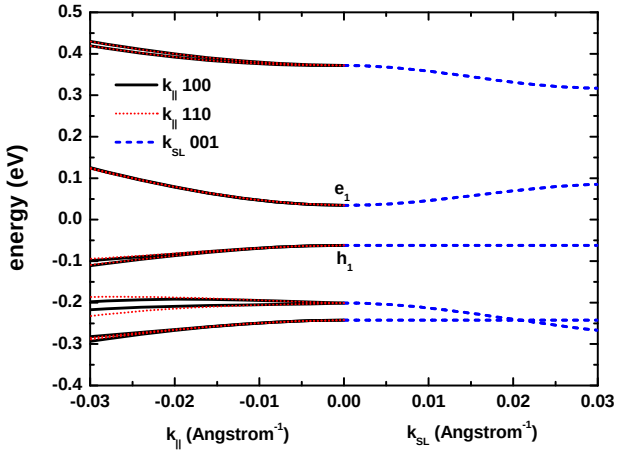


Figure 3.16.: Energy dispersion of a strained InAs–GaSb superlattice along the in-plane directions in $\mathbf{k}$ space, $\mathbf{k}_{\|}^{100}$ and $\mathbf{k}_{\|}^{110}$, and along the superlattice growth direction $\mathbf{k}_{\mathrm{SL}}$ calculated with the $\mathbf{k} \cdot \mathbf{p}$ method, where the inversion asymmetry parameter B has a nonzero value. The isotropic dispersion for small $\mathbf{k}_{\|}$ vectors is now slightly lifted. Also the twofold spin-degeneracy along $\mathbf{k}_{\|}$ is no longer present.

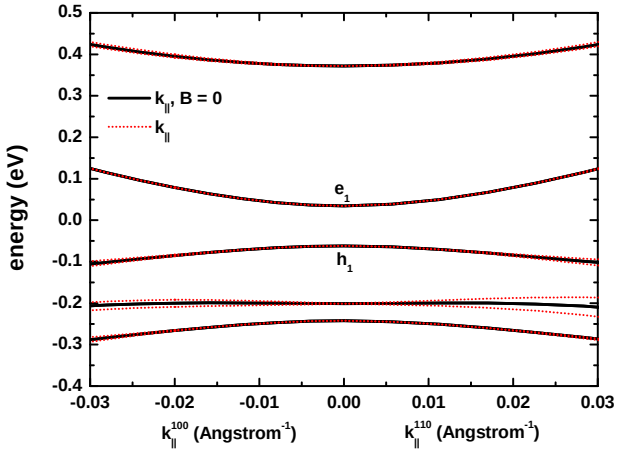


Figure 3.17.: Energy dispersion of a strained InAs–GaSb superlattice along the in-plane directions $\mathbf{k}_{\|}^{100}$ and $\mathbf{k}_{\|}^{110}$ calculated with the $\mathbf{k} \cdot \mathbf{p}$ method, where the inversion asymmetry parameter B is zero (black solid lines) and where it is nonzero (red dotted lines). In the latter case, the twofold spin degeneracy of the energy levels for nonzero $\mathbf{k}_{\|}$ is lifted due to the bulk inversion asymmetry parameter B.

4. Ballistic quantum transport using the contact block reduction (CBR) method – An introduction

This chapter is based on the work of Sabathil [Sab04] and Mamaluy et al. [MSV03] on the contact block reduction (CBR) method. I have summarized this method and presented it in an elaborate form in Ref. [BSG+09] providing several intuitive examples. This work is presented in this chapter.

The CBR method is a variant of the nonequilibrium Green's function formalism, where the local density of states is occupied using an equilibrium distribution of the carriers. It can be used to describe quantum transport in the ballistic limit very efficiently. Here, I review a numerical implementation of a charge self-consistent version of the CBR algorithm. I show in detail how to calculate the electronic properties of open quantum systems such as the transmission function, the local density of states and the carrier density. Several 1D, 2D and 3D examples are provided to illustrate the key points. The CBR method is a very powerful tool to tackle the challenge of calculating transport in the ballistic limit for 3D devices of arbitrary shape and with an arbitrary number of contacts.

4.1. Introduction

Since electronic devices have been shrinking steadily to nanometer dimensions, quantum transport is increasingly becoming a topic of interest not only to physicists but also to the electrical engineering community [Dat05]. The nonequilibrium Green's function (NEGF) formalism (e.g. Ref. [KYV+09]) provides a rigorous framework for the development of quantum device models. Here, we describe one of its implementations – the contact block reduction (CBR) method [MSV03]. It can be used to describe quantum transport in the ballistic limit very efficiently. Our aim in this article is to make the Green's function formalism in the limit of ballistic quantum transport accessible to a more general audience. Thus, a detailed description of the underlying algorithm is given and numerical examples are provided as concrete illustrations. As it is very important to perform charge self-consistent calculations, we also give details on how to solve the nonlinear system of coupled Schrödinger and Poisson equations. Interested readers should be able to reproduce these results by setting up their own computer program.

4.2. Ballistic quantum transport

A conductor shows nonohmic behavior if its dimensions are smaller than certain characteristic lengths: The mean free path and the phase-relaxation length of the electron [Dat95]. If the length of a conductor becomes shorter than the mean free path, the conductance approaches a limiting value. This classical ballistic limit has still nothing to do with quantum mechanics. Quantum mechanics does not become important until the dimensions of the conductor are smaller than the phase-relaxation length and interference-related effects come into play. In present day high-mobility semiconductor heterostructures such as modulation doped GaAs–AlGaAs heterojunctions or quantum wells, mean free paths and phase-relaxation lengths of several μm are relatively easy to obtain at low temperatures. Thus ballistic quantum transport plays an important role in many mesoscopic transport experiments.

The theoretical approach that has proven to be most useful in describing mesoscopic transport was introduced by Landauer [Lan88, Lan92] in 1988. A generalization to multiterminal devices in magnetic fields was proposed by Büttiker [Büt86, Büt88] and is generally referred to as the Landauer–Büttiker (LB) formalism. It is equivalent to the nonequilibrium Green's function formalism in the limit of no inelastic or elastic scattering. The essential idea behind the LB formalism is that the current through a ballistic conductor is determined by the probabilities of the electrons to be transmitted or reflected. The contacts of the conductor are assumed to be large electron reservoirs in equilibrium, so that each contact can be described by its own Fermi distribution with a chemical potential μ. The difference between the chemical potentials in the contacts is equal to the externally applied bias voltage. By the Landauer–Büttiker formula, these relations are expressed as follows

$$I_{\lambda\lambda'} = \frac{g_\mathrm{s} e}{h} \int T_{\lambda\lambda'}(E) \left[f(E, \mu_\lambda) - f(E, \mu_{\lambda'}) \right] \, \mathrm{d}E, \tag{4.1}$$

where $I_{\lambda\lambda'}$ is the current between contact λ and contact λ', $T_{\lambda\lambda'}(E)$ is the corresponding energy dependent transmission function between these contacts, μ_λ and $\mu_{\lambda'}$ are the chemical potentials in these contacts, E is the energy, h is Planck's constant, e is the positive elementary charge, and $g_\mathrm{s} = 2$ is the spin degeneracy of the electrons.

$$f(E, \mu_\lambda) = \frac{1}{1 + \exp\left[(E - \mu_\lambda) / (k_\mathrm{B} T) \right]} \tag{4.2}$$

is the equilibrium Fermi–Dirac distribution function inside contact λ, k_B is Boltzmann's constant and T is the temperature. Thus the Landauer–Büttiker formalism reduces the problem of calculating the ballistic current in a mesoscopic device to the determination of the transmission probabilities of an open device connected to reservoirs. We emphasize that eq. (4.1) has been simplified here. It generally involves an integration over all quantum numbers that characterize the lead states [DCVP94]. We suppressed their momentum dependence (which is, however, included in the calculations as described further below) to keep things as simple as possible and assume conservation of spin, energy E and parallel momentum. We also assume a parabolic dispersion of the bands

so that the integration over the parallel momentum can be simplified. Several numerical methods have been developed to determine the transmission coefficient for quantum devices via the scattering matrix, e.g. the transfer matrix method [Kan69, SC83], the quantum transmitting boundary method [LK90], the R-matrix method [Smr90] and the recursive Green's function method [FG97, LKBJ97]. In this article, we will describe in detail how to obtain the transmission function $T_{\lambda\lambda'}(E)$ by means of the contact block reduction (CBR) method, where the transmission is calculated from the retarded single particle Green's function. In passing, we note that the transmission function not only determines the electrical current. Also heat currents can be calculated with a Landauer formula similar to eq. (4.1) [KDL09]. Thus optimizing the thermoelectric coefficients in devices by quantum-engineering the transmission function is an interesting topic in thermoelectrics research.

4.3. The contact block reduction (CBR) method – An overview

The CBR method is a very efficient Green's function technique which has been developed by Mamaluy et al. [MSV03]. It can be used to calculate the electronic properties of open quantum systems such as the transmission function, the local density of states, and the carrier density in the ballistic limit for 1D, 2D and 3D devices of arbitrary shape and with an arbitrary number of contacts. We start with a device that is discretized in real space on N_T total grid points. It can be characterized by a corresponding Hamiltonian matrix $\mathbf{H}^0$ of size N_T. The device has no contacts and is thus termed a *closed* system. It has sharp resonant energies (eigenvalues of $\mathbf{H}^0$) and the electrons are described by wave functions (eigenfunctions of $\mathbf{H}^0$). We now add contacts to the device and divide the total number of grid points into N_C contact grid points and N_D interior device grid points ($N_\mathrm{C} + N_\mathrm{D} = N_\mathrm{T}$). Connecting the device to contacts leads to a broadening of the resonant energies: The discrete energy spectrum transforms into a continuous density of states. This is described by the broadening matrix $\mathbf{\Gamma}(E)$. It depends on energy E and has the same size as $\mathbf{H}^0$. It can directly be calculated from the self-energy $\mathbf{\Sigma}(E)$. This self-energy matrix is added to the Hamiltonian to account for the new boundary conditions due to the contacts (see Subsection 4.4.3 for details). It is non-Hermitian, thus leading to complex eigenvalues. In fact, the imaginary part of the eigenvalues is the origin of the broadening of the density of states and introduces a finite lifetime to the eigenstates. Consequently, the device wave functions leak out into the contacts (*open* device). As $\mathbf{\Sigma}$ also depends on energy, it is more convenient to look at the device from another point of view. Rather than asking for the eigenenergies of the system, it is more appropriate to ask: How does the open device *respond* to incident electrons that have a certain energy E? In the ballistic case, all observables of interest can be obtained from the retarded Green's function $\mathbf{G}^\mathrm{R}$ of the open device. It is defined as

$$\mathbf{G}^\mathrm{R}(E) = \left(E\mathbf{1} - \mathbf{H}^0 - \mathbf{\Sigma}\right)^{-1}, \tag{4.3}$$

where $\mathbf{1}$ is the identity matrix. It can be expressed in terms of the retarded Green's function $\mathbf{G}^0$ of the closed device via the Dyson equation

$$\mathbf{G}^{\mathrm{R}} = \mathbf{A}^{-1}\mathbf{G}^0 = \left(\mathbf{1} - \mathbf{G}^0\mathbf{\Sigma}\right)^{-1}\mathbf{G}^0 \tag{4.4}$$

$$\mathbf{G}^0(E) = \left(E\mathbf{1} - \mathbf{H}^0 + i0^+\right)^{-1} = \sum_n \frac{|\psi_n\rangle\langle\psi_n|}{E - E_n + i0^+}. \tag{4.5}$$

The last expression (spectral representation) shows how to write the retarded Green's function $\mathbf{G}^0$ in terms of the eigenenergies and wave functions of the closed device Hamiltonian (see Subsection 4.4.1). $|\psi_n\rangle\langle\psi_n|$ is the dyadic product where $\langle\psi_n|$ is a row vector and $|\psi_n\rangle$ is a column vector (bra–ket or Dirac notation), each of size N_{T}. In a numerical implementation of this equation, the infinitesimally small positive imaginary number $i0^+$ can be ignored if one ensures that $E \neq E_n$. Additionally, if the wave functions ψ_n are real, the retarded Green's function of the closed device is real. Thus it is identical to the advanced Green's function of the closed device. (The conjugate transpose (†) of the retarded Green's function is called the advanced Green's function $\mathbf{G}_{\mathrm{C}}^{\mathrm{A}} = \mathbf{G}_{\mathrm{C}}^{\mathrm{R}\dagger}$.) We call $\mathbf{G}^0$ just the *Green's function of the closed device* and omit $i0^+$ and the term *retarded* in the following for simplicity.

Once the self-energy matrix has been calculated (see Subsection 4.4.3), the evaluation of the retarded Green's function $\mathbf{G}^{\mathrm{R}}$ of the open device requires – in general – the inversion of a large matrix $\mathbf{A}$ whose size is proportional to the number N_{T} of total grid points of the device. Even in two spatial dimensions, this can be a quite demanding task.

The essence of the contact block reduction method consists in the decomposition of the retarded Green's function into blocks such that the transmission function of the open device can be calculated by inverting only small matrices: The retarded Green's function can be 'reduced to the contact block' $\mathbf{G}_{\mathrm{C}}^{\mathrm{R}}$. The contact block (labeled with subscript C) consists of all lead grid points that are in contact with the device. This number is orders of magnitude smaller than the number of device grid points. This explains the astonishing efficiency of this approach and makes it possible to address quantum transport in 3D devices. The CBR method has been applied to calculate the transport in 3D structures, like quantum dots [SBMV03], quantum interference devices such as a quantum logic gate [SMV04] or nano-FinFETs [VMK+08]. The latter requires us to include the Poisson equation in order to guarantee charge self-consistency (self-consistent CBR [MVS+05], see Section 4.8). The CBR method has been extended to describe systems of two interacting particles for the study of two-qubit devices [ZVB07]. It has also been extended to more sophisticated band structure models, like the $\mathbf{k}\cdot\mathbf{p}$ method in order to describe hole transport in quantum wires [MVS+05] and to tight-binding methods [RK08]. It has been integrated into the next**nano**^{3} software package [BZA+07] which is available online [www][1].

[1] Online resource: A demo that includes a Windows executable and the input files of the CBR examples presented in the figures of this chapter can be downloaded from this link: http://www.nextnano.de/customer/downloadCBR.php

In this article we describe in detail how to calculate the transmission function $T(E)$ and the local density of states $\rho(\mathbf{x}, E)$ from the Green's function matrix $\mathbf{G}^{\mathrm{R}}$ of the open system

$$\mathbf{G}^{\mathrm{R}} = \begin{pmatrix} \mathbf{G}^{\mathrm{R}}_{\mathrm{C}} & \mathbf{G}^{\mathrm{R}}_{\mathrm{CD}} \\ \mathbf{G}^{\mathrm{R}}_{\mathrm{DC}} & \mathbf{G}^{\mathrm{R}}_{\mathrm{D}} \end{pmatrix}. \tag{4.6}$$

This matrix has been subdivided into four blocks, a submatrix within the contact block (C) and another one within the interior of the device (D). The other two correlate the contact grid points to the device grid points (CD and DC). To obtain the transmission function, it is only necessary to evaluate the upper left part – the contact block

$$\mathbf{G}^{\mathrm{R}}_{\mathrm{C}} = \mathbf{A}^{-1}_{\mathrm{C}} \mathbf{G}^{0}_{\mathrm{C}}. \tag{4.7}$$

For calculating the local density of states, additionally the lower left part

$$\mathbf{G}^{\mathrm{R}}_{\mathrm{DC}} = \mathbf{G}^{0}_{\mathrm{DC}} \mathbf{B}^{-1}_{\mathrm{C}} \tag{4.8}$$

has to be evaluated. Thus for each energy E of interest, the two matrices

$$\mathbf{A}_{\mathrm{C}} = \mathbf{1}_{\mathrm{C}} - \mathbf{G}^{0}_{\mathrm{C}} \mathbf{\Sigma}_{\mathrm{C}} \tag{4.9}$$

and

$$\mathbf{B}_{\mathrm{C}} = \mathbf{1}_{\mathrm{C}} - \mathbf{\Sigma}_{\mathrm{C}} \mathbf{G}^{0}_{\mathrm{C}} \tag{4.10}$$

have to be inverted where $\mathbf{1}_{\mathrm{C}}$ is the identity matrix of dimension N_{C}. The dimension of these matrices is very small and is determined by the number N_{C} of grid points that connect the device to the contacts. For one-dimensional devices $N_{\mathrm{C}} = 2$, so both matrices are of size 2×2. In the ballistic case, the self-energy matrix $\mathbf{\Sigma}$ is nonzero only at the contact grid pints and can thus be reduced to $\mathbf{\Sigma}_{\mathrm{C}}$. This is the reason why only small parts of the Green's functions have to be evaluated. The transmission function determines the current through the device, and from the local density of states, the charge density can be derived. This is all one needs to describe quantum transport in arbitrary devices within the ballistic limit, i.e. for situations where scattering can be ignored.

4.4. The CBR method for one-dimensional devices

In this section we describe the contact block reduction (CBR) method for simple one-dimensional devices where the device geometry is assumed to be translationally invariant in the (x, y) plane. Current transport is assumed to be along the z direction. We choose the 1D case in order to highlight the main points of the CBR method, avoiding therefore to include the additional, more complicated features coming into play when one deals with two-dimensional and three-dimensional devices described in Section 4.5. We try to avoid reproducing the equations and the arguments of the original CBR papers [MSV03, MVS+05] and adopt the more straightforward approach to focus specifically on the aspects with respect to a numerical implementation.

In a one-dimensional device one can only have two leads (i.e. contacts) in total ($L = 2$). These leads are located at the leftmost and rightmost boundary points of the device and each lead λ contains exactly one grid point ($N_\lambda = 1$) that connects the lead to the device, i.e. the total number of (relevant) lead grid points is thus equal to $N_\mathrm{C} = \sum_{\lambda=1}^{L} N_\lambda = 2$. This simplifies the CBR method substantially because the dimension of the CBR contact matrices is exactly equal to $N_\mathrm{C} = 2$. This means that for the calculation of the transmission coefficient $T(E)$ (see eq. (4.1)), for each energy E only a small square matrix of size $N_\mathrm{C} = 2$ has to be inverted. A further simplification is that each lead has only one mode. In a 2D or 3D simulation, each lead consists of several lead grid points connected to the device ($N_\lambda > 1$). The number of lead grid points corresponds to the number of lead modes (see Section 4.5), i.e. each lead has N_λ modes. In a 1D simulation, the CBR algorithm is then implemented as follows:

4.4.1. Energy levels and wave functions of the device Hamiltonian (closed system)

First, we calculate the energy levels and the wave functions of the device Hamiltonian without taking the leads into account. This Hamiltonian $\mathbf{H}^0$ is then identical to the Hamiltonian of the closed system. We use a standard approach to solve the Schrödinger equation, namely the envelope function approximation assuming a parabolic dispersion (single-band effective mass equation).

The Schrödinger equation for a semiconductor structure grown along the z direction and homogeneous along the x and y directions reads

$$\mathbf{H}^0_{\mathbf{k}_\parallel} \Psi_n\left(z, \mathbf{k}_\parallel\right) = E_n\left(\mathbf{k}_\parallel\right) \Psi_n\left(z, \mathbf{k}_\parallel\right). \tag{4.11}$$

The wave function $\Psi_n\left(z, \mathbf{k}_\parallel\right)$ can by factorized into a solution $\psi_n\left(z, \mathbf{k}_\parallel\right)$ along the z direction, and a plane wave $e^{i\mathbf{k}_\parallel \cdot \mathbf{x}}$ in the (x, y) plane

$$\Psi_n\left(z, \mathbf{k}_\parallel\right) = \psi_n\left(z, \mathbf{k}_\parallel\right) e^{i\mathbf{k}_\parallel \cdot \mathbf{x}}. \tag{4.12}$$

In the following we ignore the dependence of $\psi_n\left(z, \mathbf{k}_\parallel\right)$ on the parallel momentum $\mathbf{k}_\parallel$. Then the envelope functions $\psi_n(z)$ of the n_T quantized states are obtained as the solutions of the one-dimensional Schrödinger equation ($n = 1, ..., n_\mathrm{T}$ where $n_\mathrm{T} = N_\mathrm{T}$):

$$\mathbf{H}^0 \psi_n(z) = E_n \psi_n(z) \tag{4.13}$$

$$\left[-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m_\perp(z)}\frac{\partial}{\partial z}\right) + V(z)\right]\psi_n(z) = E_n \psi_n(z) \tag{4.14}$$

$m_\perp(z)$ is the effective mass tensor component along the z direction, $\hbar$ is Planck's constant divided by 2π, $V(z) = E_\mathrm{c}(z) = E_{\mathrm{c},0}(z) - e\phi(z)$ is the spatially varying potential energy (conduction band edge profile), $E_{\mathrm{c},0}(z)$ represents the conduction band edge profile of the particular band of interest including band offsets at material interfaces and $\phi(z)$ is the electrostatic potential which is obtained from solving Poisson's equation (see Subsection 4.8.1). It includes the external bias potential and the internal potential resulting from mobile charge carriers and ionized impurities.

We discretize this equation with a finite differences method on a uniform grid using Neumann boundary conditions at the left and right device boundaries. At these points the device is in contact to the leads, once they are added to form the open system. It has been found that Neumann boundary conditions at the contact grid points are much better suited for the CBR method than Dirichlet boundary conditions [MSV03]. The discretized sparse, square and Hermitian (in most cases even real and symmetric) Hamiltonian matrix of size equal to the number of total device grid points N_T has to be diagonalized numerically to yield the eigenvalues and eigenvectors.

The eigenenergies E_n correspond to the energies of the electron along the z direction. The total energy of the electron includes the parallel momentum of the electron due to $\mathbf{k}_\parallel = (k_x, k_y)$

$$E_n\left(\mathbf{k}_\parallel\right) = E_n + \frac{\hbar^2}{2m_\parallel}\mathbf{k}_\parallel^2, \tag{4.15}$$

where $m_\parallel$ derives from the mass tensor components in the (x, y) plane. For more detailed information on how to solve eq. (4.14) numerically, we refer to e.g. Ref. [TSCH90].

The one-dimensional envelope functions ψ_n are usually normalized to 1

$$\int \psi_n^*(z)\psi_n(z)\ \mathrm{d}z = \sum_{i=1}^{N_\mathrm{T}} \psi_{n,i}^*\psi_{n,i}\Delta_i = 1, \tag{4.16}$$

where $\psi_{n,i}$ is the amplitude of the wave function at grid point i, and Δ_i the corresponding grid spacing along the z direction. If the latter has units of [nm], then the wave functions ψ_n have units of $\left[\mathrm{nm}^{-1/2}\right]$. In principle, the wave functions could have been calculated using a nonuniform grid spacing. However, in the following sections we assume that the grid spacing Δ_i is homogeneous for all grid points. This allows us to renormalize the wave functions so that they become dimensionless. This is achieved by dividing ψ_n by the norm $1/\sqrt{\Delta}$. Then the normalization reads

$$\sum_{i=1}^{N_\mathrm{T}} \psi_{n,i}^*\psi_{n,i} = 1. \tag{4.17}$$

Incomplete set of eigenstates We want to emphasize that the actual number n_α of eigenvalues and wave functions needed to get meaningful results within the CBR method can be much smaller than the total number n_T of eigenfunctions of the Hamiltonian matrix. The energy E_{n_α} of the highest eigenvector taken into account is the cutoff energy. It should be significantly above the energy interval of interest in order to get reliable results (see Fig. 4.4 and Fig. 4.6). For a 2D and 3D simulation, using such an incomplete set of eigenstates will drastically improve the computational performance as only n_α eigenstates have to be calculated ($n_\alpha \approx 10\%$ of n_T). This fact makes it attractive to use fast, iterative solvers [LSY98] for calculating only a small number of eigenstates of these sparse matrices. In 1D, where it is not a computational challenge to calculate all eigenvalues of the spectrum, exact solvers [LAP] might be preferable that either return all or only the requested number of eigenvalues (or eigenvalues within the specified energy interval of interest).

4.4.2. Projection of device eigenfunctions onto lead modes

This part is very easy for a one-dimensional simulation where for each of the two leads only one lead mode exists. One simply has to store – for each eigenvalue n – the values of the wave functions $\psi_{n,i}$ at the leftmost grid point ($i = 1$) and at the rightmost grid point ($i = N_\mathrm{T}$)

$$\text{Lead 1}: \ \chi_n^{\lambda=1} = \psi_{n,1} \qquad \text{(left boundary)} \tag{4.18}$$

$$\text{Lead 2}: \ \chi_n^{\lambda=2} = \psi_{n,N_\mathrm{T}} \ \text{(right boundary)}. \tag{4.19}$$

For each eigenvalue n these projected eigenvector amplitudes χ_n^λ are stored in a vector of size $N_\mathrm{C} = 2$

$$\chi_n = \begin{pmatrix} \chi_n^{\lambda=1} \\ \chi_n^{\lambda=2} \end{pmatrix}. \tag{4.20}$$

4.4.3. Setup energy interval and calculate properties for each energy E_i

We are interested in the transmission coefficient $T_{12}(E)$ from lead $\lambda = 1$ (left contact) to lead $\lambda = 2$ (right contact) for all energy values E in the energy interval of interest ($E_\mathrm{min} < E < E_\mathrm{max}$). To do this, we divide this energy interval into N_E energy grid points and calculate for each the transmission coefficient $T_{12}(E_i)$ from lead 1 to lead 2 for the energy value $E_i = E_\mathrm{min} + (i-1)\,\Delta_E$ where $\Delta_E = \frac{(E_\mathrm{max} - E_\mathrm{min})}{N_E - 1}$ is the energy grid spacing and $i = [1, ..., N_E]$. For each energy E_i the following matrices have to be calculated:

- self-energy matrix $\mathbf{\Sigma}_\mathrm{C}(E_i)$
- broadening matrix $\mathbf{\Gamma}_\mathrm{C}(E_i)$
- Green's function $\mathbf{G}^0_\mathrm{C}(E_i)$ of the closed device
- retarded Green's function $\mathbf{G}^\mathrm{R}_\mathrm{C}(E_i)$ of the open device

For the latter, a square matrix of dimension N_C has to be inverted (for each energy E_i). The subscript C (contact) indicates that all quantities are reduced contact block matrices of size N_C, i.e. relatively small matrices that have to be evaluated only at the boundary points where the device overlaps with the contact grid points. In a 1D simulation, $N_\mathrm{C} = 2$, so that only 2×2 matrices occur. The energy E_i corresponds to the energy E_z of the electron along the z direction because in 1D the transmission coefficient is a function of the energy E_z only: $T(E_i) = T(E_z)$. The energy due to the parallel momentum of the electron does not have to be considered for calculating T. However, one should keep in mind that the total energy of the electron is given by

$$E_\mathrm{total} = E_z + \frac{\hbar^2}{2m_\parallel}\mathbf{k}_\parallel^2, \tag{4.21}$$

which becomes relevant when calculating the density and the current.

Self-energy matrix $\mathbf{\Sigma}_{\mathrm{C}}$

$\mathbf{\Sigma}_{\mathrm{C}}$ is the contact self-energy matrix which represents the coupling of the device to the leads. The self-energy matrix in a real space representation is nonzero only at the boundary points of the device which are in contact with the leads. In a mode space representation (see Subsection 4.5.2) the self-energy matrix $\mathbf{\Sigma}$ is a diagonal matrix. In 1D the contact self-energy matrix $\mathbf{\Sigma}_{\mathrm{C}}$ has only two nonzero entries on the diagonal ($\mathbf{\Sigma}_{\lambda=1}$ and $\mathbf{\Sigma}_{\lambda=2}$)

$$\mathbf{\Sigma}_{\mathrm{C}} = \begin{pmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \mathbf{\Sigma}_2 \end{pmatrix}. \tag{4.22}$$

We assume that each lead is represented by a semi-infinite one-dimensional wire described by a one-band effective mass Hamiltonian. The potential energy E_{c}^{λ} of this contact Hamiltonian is equal to the conduction band edge energy of the corresponding grid point at the left or right device boundary. Then the contact self-energy $\mathbf{\Sigma}_{\lambda}$ for lead λ is given by [Dat05]

$$\mathbf{\Sigma}_{\lambda} = -t \exp\left(ik^{\lambda}\Delta\right), \tag{4.23}$$

where t is the kinetic coupling matrix element (also called constant intersite coupling element). It is defined as

$$t = \frac{\hbar^2}{2m}\frac{1}{\Delta^2}, \tag{4.24}$$

where m is the effective electron mass of the contact, and Δ is the grid spacing of the contact grid point along the propagation direction z. The wave vector $k^{\lambda}(E_i)$ of lead λ has to be calculated for each energy E_i from the lead dispersion $E(k^{\lambda})$. The dispersion of a discrete lattice is given by

$$E(k^{\lambda}) = E^{\lambda} + 2t\left(1 - \cos\left(k^{\lambda}\Delta\right)\right), \tag{4.25}$$

where we assume the lead to be discretized with the same grid spacing Δ. Thus the corresponding wave vector k^{λ} is obtained as follows

$$k^{\lambda}(E_i) = \frac{1}{\Delta}\arccos\left(\frac{E_i - E^{\lambda}}{2t} - 1\right). \tag{4.26}$$

$\arccos(x)$ is the inverse function of the trigonometric $\cos(x)$ function which must be expressed using the complex logarithm

$$\arccos(x) = -i\ln\left(x + i\sqrt{1 - x^2}\right) \tag{4.27}$$

in order to allow for complex k vectors. For real wave vectors, the self-energy $\mathbf{\Sigma}_{\lambda}$ corresponds to a traveling plane wave (eq. (4.23)) with a particular energy. The response of the open system to an incident electron wave tells us if this electron wave will be reflected or transmitted. Complex wave vectors, on the other hand, give rise to exponentially rising (unphysical) or decaying waves. Here, we only consider the decaying evanescent waves. In 1D, the conduction band edge energy E_{c}^{λ} at the corresponding lead λ has to

be taken for the energy E^λ. In general, the relation for the wave numbers k will differ at each contact. In 1D, there is only one mode for each lead, so only one k vector for each lead has to be calculated (for each energy). Consequently, the contact self-energy $\mathbf{\Sigma}_\lambda$ is a scalar for each lead but in general it is a matrix whose size is determined by the number of contact grid points of this lead (or lead modes taken into account, respectively). In this work, the concept of self-energy only describes the coupling of the device to the leads. However, this concept is far more general and can be used to describe all kinds of interactions, e.g. scattering processes that can be included in more advanced NEGF algorithms [KYV+09].

Broadening matrix $\mathbf{\Gamma}_\mathrm{C}$

The broadening matrix $\mathbf{\Gamma}_\mathrm{C}$ is the anti-Hermitian part of $\mathbf{\Sigma}_\mathrm{C}$ and corresponds to the broadened density of states in the device. It has units of energy and is given by

$$\mathbf{\Gamma}_\mathrm{C} = i\left(\mathbf{\Sigma}_\mathrm{C} - \mathbf{\Sigma}_\mathrm{C}^\dagger\right). \tag{4.28}$$

The eigenstates of the closed system Hamiltonian correspond to sharp energy levels, and thus they have an infinite lifetime: An electron in one of these states will stay there forever. In contrast, the broadening matrix $\mathbf{\Gamma}_\mathrm{C}$ describes the leakage of the eigenstates into the contacts. Consequently, this will lead to a finite lifetime of the electronic states in the device.

Green's function $\mathbf{G}_\mathrm{C}^0$ of the closed device

The reduced contact block matrix $\mathbf{G}_\mathrm{C}^0(E_i)$ can be written in terms of the projected wave functions χ_n of the decoupled device Hamiltonian $\mathbf{H}^0$ at the contact grid points

$$\mathbf{G}_\mathrm{C}^0(E_i) = \sum_{n=1}^{n_\alpha} \frac{|\psi_{n,\mathrm{C}}\rangle\langle\psi_{n,\mathrm{C}}|}{E_i - E_n} = \sum_{n=1}^{n_\alpha} \frac{\chi_n \chi_n^\mathrm{T}}{E_i - E_n}. \tag{4.29}$$

Here we use the spectral representation in terms of the wave function amplitudes $\psi_{n,\mathrm{C}}$ of the closed device Hamiltonian at the contacts to calculate the Green's function $\mathbf{G}_\mathrm{C}^0$. $\chi_n\chi_n^\mathrm{T}$ is the dyadic product where χ_n is a column vector and χ_n^T its transpose, i.e. a row vector, each of size N_C containing the projection of the wave function amplitude onto the two lead grid points (see Subsection 4.4.2). In 1D this dyadic product leads to a square matrix of dimension $N_\mathrm{C} = 2$. Only for the exact solution, one has to take into account all n_T eigenstates ($n_\alpha = n_\mathrm{T}$). For 2D and 3D simulations, n_α is typically chosen to be much smaller ($\approx 10\%$ of all eigenstates), making use of an incomplete set of eigenstates. To guarantee an optimal use of the CBR method, the value of n_α should be chosen as small as possible to minimize computational effort. However, one has to ensure that it is still large enough in order to get meaningful results for the energy interval of interest (see Subsection 4.4.1).

The matrix $\mathbf{G}_\mathrm{C}^0$ can further be understood by decomposing it into submatrices

$$\mathbf{G}_\mathrm{C}^0 = \begin{pmatrix} \mathbf{G}_{\lambda\lambda}^0 & \mathbf{G}_{\lambda\lambda'}^0 \\ \mathbf{G}_{\lambda'\lambda}^0 & \mathbf{G}_{\lambda'\lambda'}^0 \end{pmatrix} = \begin{pmatrix} \mathbf{G}_{11}^0 & \mathbf{G}_{12}^0 \\ \mathbf{G}_{21}^0 & \mathbf{G}_{22}^0 \end{pmatrix}. \tag{4.30}$$

The submatrix $\mathbf{G}^0_{\lambda\lambda'}$ couples lead λ to lead λ'. In 1D this submatrix is a scalar because χ^λ_n is a scalar

$$\mathbf{G}^0_{\lambda\lambda'}(E_i) = \sum_{n=1}^{n_\alpha} \frac{|\psi^\lambda_{n,\mathrm{C}}\rangle\langle\psi^{\lambda'}_{n,\mathrm{C}}|}{E_i - E_n} = \sum_{n=1}^{n_\alpha} \frac{\chi^\lambda_n \chi^{\lambda'}_n}{E_i - E_n}. \tag{4.31}$$

Retarded Green's function $\mathbf{G}^\mathrm{R}_\mathrm{C}$ of the open device

In order to calculate the transmission coefficient, we first have to evaluate the retarded Green's function $\mathbf{G}^\mathrm{R}_\mathrm{C}$ within the contact region from the Dyson equation

$$\mathbf{G}^\mathrm{R}_\mathrm{C} = \mathbf{A}^{-1}_\mathrm{C}\mathbf{G}^0_\mathrm{C} \tag{4.32}$$

$$\mathbf{A}_\mathrm{C} = \mathbf{1}_\mathrm{C} - \mathbf{G}^0_\mathrm{C}\mathbf{\Sigma}_\mathrm{C}, \tag{4.33}$$

where $\mathbf{1}_\mathrm{C}$ is the identity matrix and $\mathbf{G}^0_\mathrm{C}\mathbf{\Sigma}_\mathrm{C}$ is a simple matrix multiplication. $\mathbf{G}^\mathrm{R}_\mathrm{C}$ is a small submatrix of size 2×2 of the open device's retarded Green's function $\mathbf{G}^\mathrm{R}$ within the contact regions (see eq. (4.6)). $\mathbf{G}^\mathrm{R}$ has the size of the total number of grid points N_T and is thus a very large matrix for 2D and 3D devices. The direct evaluation of the retarded Green's function requires the inversion of a large matrix of dimension N_T which is practically impossible for a 3D device, and can be quite demanding even in two spatial dimensions. The essence of the CBR method consists in realizing that for the calculation of the transmission function, only the small part $\mathbf{G}^\mathrm{R}_\mathrm{C}$ is needed. The determination of this small submatrix from $\mathbf{G}^0$ and $\mathbf{\Sigma}$ actually requires only the inversion of a matrix that is proportional to the number of grid points N_C that connect the device with the leads.

The inversion of the matrix $\mathbf{A}_\mathrm{C}$ to obtain $\mathbf{A}^{-1}_\mathrm{C}$ is the central part of the CBR algorithm because in a 2D or 3D simulation, most of the CPU time is consumed here. The inversion can be performed by a standard inversion routine from a numerical library (e.g. LAPACK routine ZGESV [LAP] which is also available from precompiled libraries that make efficient use of multicore processor architectures). For a matrix of dimension N_C, this usually requires of the order of $(N_\mathrm{C})^{2.8}$ to $(N_\mathrm{C})^3$ operations. Luckily, N_C is generally very small because the number of contact grid points is much smaller than the number of device grid points.

Transmission coefficient

Finally, we calculate for each energy the transmission coefficient $T_{\lambda\lambda'}(E_i)$ from the broadening matrix $\mathbf{\Gamma}_\mathrm{C}$ and the retarded Green's function $\mathbf{G}^\mathrm{R}_\mathrm{C}$ within the contact region

$$T_{\lambda\lambda'}(E_i) = \mathrm{Tr}\left(\mathbf{\Gamma}^\lambda_\mathrm{C}\mathbf{G}^\mathrm{R}_\mathrm{C}\mathbf{\Gamma}^{\lambda'}_\mathrm{C}\mathbf{G}^{\mathrm{R}\dagger}_\mathrm{C}\right) \qquad (\lambda \neq \lambda'), \tag{4.34}$$

where † indicates the conjugate transpose. The three matrix multiplications only have to be performed for the relevant elements that contribute to the trace of the square matrix of dimension N_C. The elements of the small matrix $\mathbf{G}^\mathrm{R}_\mathrm{C}$ completely determine the transmission function from lead $\lambda = 1$ to lead $\lambda' = 2$. $\mathbf{\Gamma}^\lambda_\mathrm{C}$ is the broadening matrix

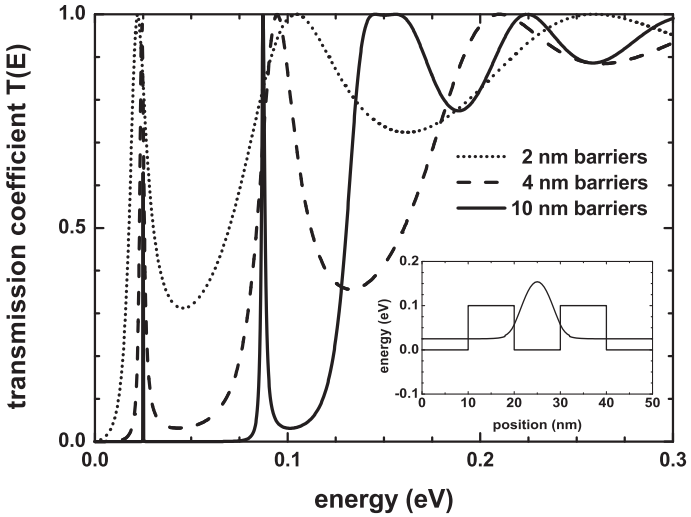


Figure 4.1.: Calculated transmission coefficient $T(E)$ as a function of energy for a double barrier structure with varying barrier widths of 2 nm, 4 nm and 10 nm (barrier height 100 meV, barrier separation 10 nm). At 25 meV there is a peak where the double barrier becomes transparent, i.e. $T(E) = 1$. This is exactly the energy that matches the resonant state in the well. The inset shows the conduction band edge profile and the probability density of this quasibound resonant state for the case of 10 nm barrier widths.

for lead λ defined analogously to eq. (4.28). It is nonzero only at the contact points of the relevant lead. In the basis we employ here, it is a diagonal matrix. In fact, for a 1D simulation only one element of this matrix is nonzero and the calculation of the transmission coefficient involves the multiplication of four scalars, two of them are due to the nonzero entries of the broadening matrices of the leads, and the other two originate from the off-diagonal elements of the reduced matrix $\mathbf{G}_{\mathrm{C}}^{\mathrm{R}}$.

4.4.4. Transmission function of a double barrier structure (1D example)

As a simple intuitive example we present in Fig. 4.1 the calculated transmission coefficient $T(E)$ as a function of energy for a double barrier structure with varying barrier widths of 2 nm, 4 nm and 10 nm (barrier height 100 meV, barrier separation 10 nm, effective mass $m = 0.067m_0$, grid spacing 0.5 nm, device length 50 nm). At 25 meV there is a peak where the double barrier becomes transparent, i.e. $T(E) = 1$. This is exactly the energy that matches the resonant state in the well. The inset shows the conduction band edge profile and the probability density of this quasibound resonant state for the case of 10 nm barrier widths where the resonant state hardly couples to the two leads. In the opposite case of strong coupling of this resonant state to the leads (2 nm barrier widths), the local density of states (LDOS) $\rho(z, E)$ around this resonant state broadens,

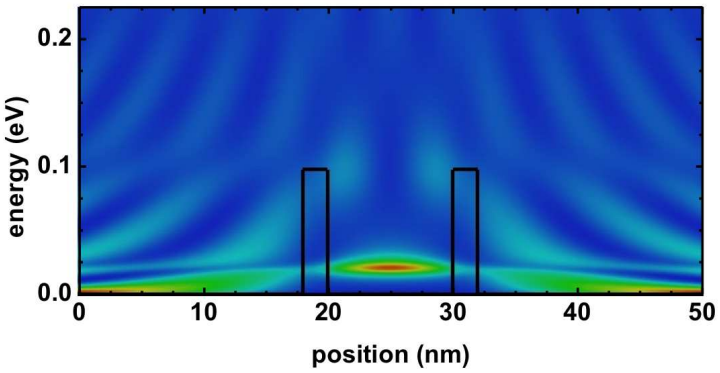


Figure 4.2.: Calculated local density of states $\rho(z, E)$ for a double barrier structure (barrier widths 2 nm, barrier height 100 meV, barrier separation 10 nm). The conduction band edge profile is indicated by the thick solid line. The resonant state inside the double barrier is very broad with respect to energy because it couples strongly to the leads at the left and right boundaries. This is in contrast to the situation for the 10 nm barriers (not shown) where due to the large barrier widths the resonant state is quasibound, i.e. with a very sharp and high density of states at the resonance energy because of the very weak coupling to the contacts. Red (blue) color indicates high (low) density of states.

leading to a broadening of the peaks in the transmission coefficient. This is shown in Fig. 4.2 where the LDOS is plotted as a function of position and energy for the 2 nm case. The red (blue) color indicates high (low) density of states. This is in contrast to the situation for the 10 nm barriers (not shown) where due to the large barrier widths the resonant state is quasibound, i.e. with a very sharp and high density of states at the resonance energy because of the very weak coupling to the contacts. If the energy grid is not fine enough, very sharp resonances can be missed in a numerical calculation. This is the reason why we used an energy grid spacing of 0.5 meV. However, this grid spacing is still not fine enough to get perfect transmission ($T = 1$) for the first peak of the 10 nm barrier structure. Only if the energy grid point exactly matches the resonance energy, the peak would be well resolved. As it is very instructive to investigate the local density of states in different parts of the device, we will show in Section 4.6 how to calculate it with the CBR method.

The calculated density of states (DOS) for the 2 nm, 4 nm and 10 nm double barrier structures is shown in Fig. 4.3. The DOS corresponds to the LDOS integrated over position. The first peak in the DOS for the 10 nm barrier structure differs substantially from the other two structures because it is extremely sharp and high. It is actually much higher than the figure suggests because its maximum is not included on this scale. The second peak in the DOS at 87 meV due to the second confined well state is only visible for the 10 nm structure. This is consistent to the transmission coefficient (see Fig. 4.1)

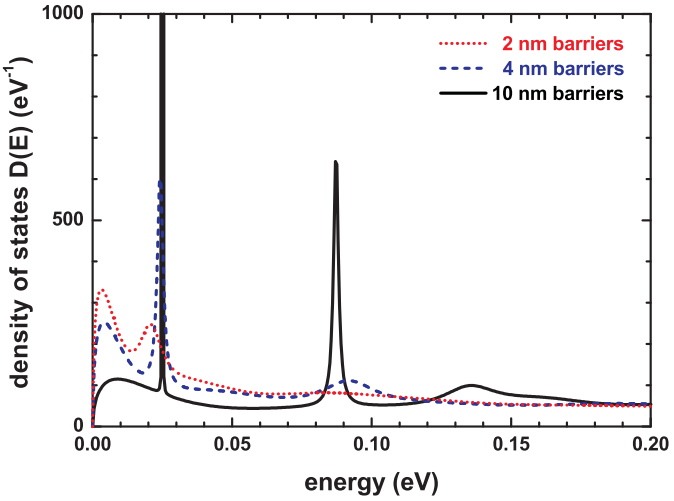


Figure 4.3.: Calculated density of states $D(E)$ for the double barrier structures. The first peak in the DOS for the 10 nm barrier structure differs substantially from the other two structures because it is extremely sharp and high. The second peak in the DOS at 87 meV due to the second confined well state is only visible for the 10 nm structure. This is consistent to the transmission coefficient (see Fig. 4.1) which shows a sharp maximum only for the 10 nm structure at this energy.

which shows a sharp maximum only for the 10 nm structure at this energy.

Figure 4.4 shows the calculated transmission coefficient $T(E)$ of the 2 nm double barrier structure highlighting the CBR feature of using an incomplete set of eigenstates (10%, 40% and 100% of the eigenstates of the closed device Hamiltonian). Even if only 10% of the eigenstates are used, the first resonance can nicely be reproduced. The cutoff energy in this case is at 180 meV which explains the sudden drop in $T(E)$ for energies exceeding this value. Using 40% of the eigenstates, the main features in the energy interval of interest can be reproduced very well because only the low-energy part of the retarded Green's function is relevant for the transmission function near the band edge.

In Fig. 4.1, Fig. 4.2, Fig. 4.3 and Fig. 4.4, the energy E actually refers to the energy E_z of the electron along the z direction.

4.5. The CBR method for two- and three-dimensional devices

This section has the same structure as the one for one-dimensional devices. We only mention the differences with respect to the 1D devices. The most important aspect is, that now we have to deal with lead modes. For 2D devices the contacts are one-dimensional lines with one-dimensional eigenfunctions. An example is shown in Fig. 4.7 that is further discussed in Subsection 4.5.4. For 3D devices the contacts are two-dimensional surfaces

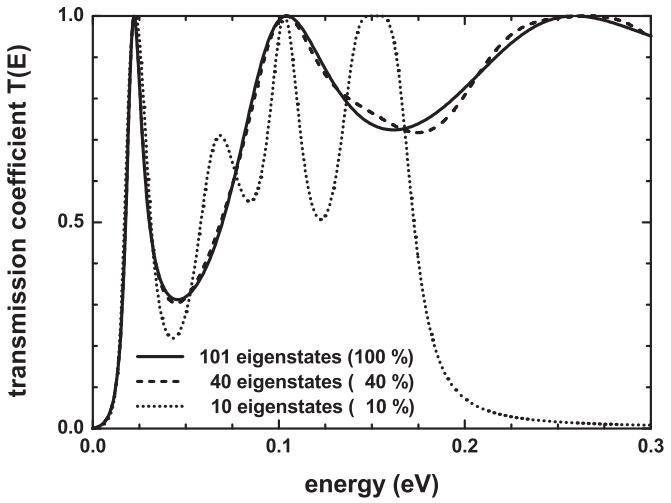


Figure 4.4.: Calculated transmission coefficient $T(E)$ of a double barrier structure (barrier widths of 2 nm) showing the CBR feature of using an incomplete set of eigenstates (10%, 40% and 100% of the eigenstates of the closed device Hamiltonian). Even if only 10% of the eigenstates are used, the first resonance can nicely be reproduced. The cutoff energy in this case is at 180 meV which explains the sudden drop in $T(E)$ for energies exceeding this value. Using 40% of the eigenstates, the main features in the energy interval of interest can be reproduced very well.

leading to two-dimensional lead eigenfunctions (Subsection 4.5.5). To obtain these lead eigenfunctions χ_m^λ a corresponding one- or two-dimensional Schrödinger equation has to be solved for each lead. The 1D Schrödinger equation is identical to eq. (4.13), and the normalization of the wave functions has to be done consistently to the device wave functions (eq. (4.16), eq. (4.17)). The dimension of the contact Hamiltonian matrix is given by the number N_λ of contact grid points connecting this lead to the device. The total number of modes m^λ of this lead is then also equal to N_λ.

4.5.1. Energy levels and wave functions of the device Hamiltonian (closed system)

For 2D and 3D devices, the corresponding two-dimensional and three-dimensional Schrödinger equations are solved for the closed system. At the device boundary grid points where the device is in contact to the leads, Neumann boundary conditions are employed along the propagation direction, i.e. perpendicular to the lead line (2D device) or lead surface (3D device). For all other device boundary grid points that are not connected to leads, Dirichlet boundary conditions are taken. Usually, only a small number n_α of the total number n_T of eigenvectors has to be calculated.

4.5.2. Projection of device eigenfunctions onto lead modes

In general, the vector χ_n of eq. (4.20) has now the following structure

$$\chi_n = \begin{pmatrix} \chi^{\lambda=1}_{n,m=1} \\ \cdots \\ \chi^{\lambda=1}_{n,m=m^1} \\ \cdots \\ \chi^{\lambda=L}_{n,m=1} \\ \cdots \\ \chi^{\lambda=L}_{n,m=m^L} \end{pmatrix}, \tag{4.35}$$

and takes into account that for each lead λ ($\lambda = 1, ..., L$) several lead modes m ($m = 1, ..., m^\lambda$) exist. The components of the vector χ_n are calculated by projecting the parts of the device eigenvectors $\psi_{n,\mathrm{C}}$ (real space representation) that are in contact to the leads into the basis of the orthogonal lead eigenfunctions (mode space representation). For each eigenvalue n, the amplitude of the wave function $\psi^\lambda_{n,i}$ at the contact grid point i is projected onto the amplitude of the lead eigenfunction $\chi^\lambda_{m,i}$ of mode m at this contact grid point

$$\chi^\lambda_{n,m} = \sum_i \langle \psi^\lambda_{n,i} \mid \chi^\lambda_{m,i} \rangle. \tag{4.36}$$

The sum runs over all contact grid points i of the relevant lead λ.

Incomplete set of lead modes Within the mode space basis the self-energy matrix $\mathbf{\Sigma}_\mathrm{C}$ is diagonal and can be truncated at the cutoff energy. We want to emphasize that the actual number m^λ_α of lead eigenvalues and lead wave functions needed to get meaningful results within the CBR method can be much smaller than the total number m^λ of lead modes of this lead Hamiltonian matrix. Neglecting nonpropagating high energy modes reduces the size of the contact block matrices that have to be inverted for each energy. The new size is then given by $N_{\mathrm{C,m}}$ which is also the size of the vector χ_n that now only takes into account the modes up to m^λ_α for each lead. For an exact solution all modes m^λ have to be included. The energy $E_{m^\lambda_\alpha}$ of the highest lead eigenvector taken into account is the cutoff energy for this lead. It should be significantly above the energy interval of interest in order to get reliable results. This lead mode cutoff energy should have about the same value as for the expansion of $\mathbf{G}^0_\mathrm{C}$ (see eq. (4.29)). For a 2D and 3D simulation, using such an incomplete set of lead modes will significantly improve the computational performance. For further details on the transformation into the subspace of the propagating lead modes for a 2D or 3D device, we refer to Section V. “Mode space reduction in single-band case” of Ref. [MVS+05].

4.5.3. Setup energy interval and calculate properties for each energy E_i

For a 3D simulation, the energy E_i corresponds to the total electron energy whereas for a 2D simulation $E_i = E_{x,y}$ with $E_{x,y}$ being the energy of the electron in the (x, y) plane

$$E_{\text{total}} = E_{x,y} + \frac{\hbar^2}{2m_{\parallel}} k_{\parallel}^2. \tag{4.37}$$

Here, we assume the device to be translationally invariant along the z direction ($k_{\parallel} = k_z$, $m_{\parallel} = m_z$).

Self-energy matrix $\mathbf{\Sigma}_{\text{C}}$

Within the basis of the orthogonal lead eigenfunctions (mode space representation) the self-energy matrix $\mathbf{\Sigma}_{\text{C}}$ is diagonal

$$\mathbf{\Sigma}_{\text{C}} = \begin{pmatrix} \mathbf{\Sigma}_{\lambda=1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \mathbf{\Sigma}_{\lambda=L} \end{pmatrix}. \tag{4.38}$$

For each lead a small diagonal submatrix $\mathbf{\Sigma}_{\lambda}$ of dimension m_{α}^{λ} has to be calculated. Its components are the contact self-energies $\mathbf{\Sigma}_{\lambda}^{m}$ for each mode m ($m = 1, ..., m_{\alpha}^{\lambda}$) of the relevant lead

$$\mathbf{\Sigma}_{\lambda}^{m} = -t \exp\left(i k_m^{\lambda} \Delta\right). \tag{4.39}$$

Therefore a wave vector k_m^{λ} for each lead and for each transverse mode m has to be calculated (for each energy E_i)

$$k_m^{\lambda}(E_i) = \frac{1}{\Delta} \arccos\left(\frac{E_i - E_m^{\lambda}}{2t} - 1\right). \tag{4.40}$$

Obviously, the propagation direction can now be along the x, y or z direction, depending on the orientation of the lead with respect to the device. Then for the contact boundary grid point the corresponding mass tensor component and the grid spacing Δ along the appropriate propagation direction has to be taken for t. Eq. (4.40) can be derived from the dispersion of a discrete lattice

$$E(k_m^{\lambda}) = E_m^{\lambda} + 2t\left(1 - \cos\left(k_m^{\lambda}\Delta\right)\right), \tag{4.41}$$

where E_m^{λ} is the eigenenergy of the m^{th} mode of lead λ.

4.5.4. Transmission function of a 2D structure with several barriers (2D example)

As a simple 2D illustration we take the same example as presented in Ref. [MSV03]. The structure consists of three leads with a Gaussian shaped barrier of height 1.0 eV in the middle and a double barrier in the upper part of the device with a height of 0.4 eV. The

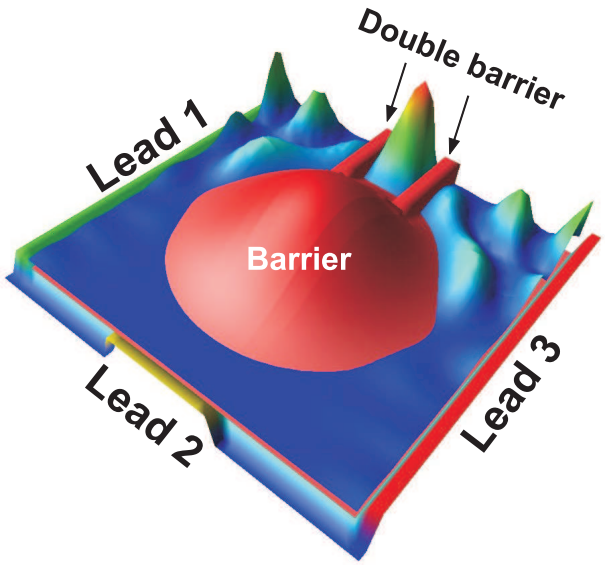


Figure 4.5.: Conduction band profile (barrier of 2D Gaussian shape and a double barrier of height 0.4 eV) of a 2D device that is connected to three leads. Indicated is the square of the wave function of the 26$^{\text{th}}$ eigenstate which is a resonance state of the device where the transmission coefficient $T_{13}(E)$ between lead 1 and lead 3 shows a local maximum at around 0.18 eV (see Fig. 4.6).

device has a width of 20 nm and is discretized with 41 grid points in each direction leading to a Hamiltonian matrix of dimension $N_{\text{T}} = 1681$ (grid spacing 0.5 nm). For further details we refer to the original publication [MSV03]. Figure 4.5 shows the conduction band edge profile and the square of the wave function of the 26$^{\text{th}}$ eigenstate which is a resonance state of the device where the transmission coefficient $T_{13}(E)$ between lead 1 and lead 3 shows a local maximum at around 0.18 eV (see Fig. 4.6). This corresponds to resonant tunneling in the upper path where the electron tunnels through the double barrier. The first peak at 0.11 eV in the calculated transmission coefficient is not due to a resonance of the double barrier – as one might first be tempted to guess. It is related to the electron travelling the lower path around the Gaussian shaped barrier. Such information can be obtained by the visualization of the relevant wave functions or local density of states at this energy (not shown). This example demonstrates that even for very simple structures, it is vital to have access to calculated quantum mechanical properties in order to characterize the peaks correctly.

Figure 4.6 shows that using an incomplete set of eigenstates of only 7% (118 of 1681) of the 2D device Hamiltonian of the closed system is sufficient to calculate the transmission coefficient up to energies of 0.4 eV. In 1D devices the transmission function cannot exceed the value of 1. For 2D and 3D devices the maximum value of the transmission function is obtained if each of the m^{λ} lead modes in one lead transmits perfectly to the other lead. So in our example where the leads 1 and 3 each have 41 modes, the maximum of the

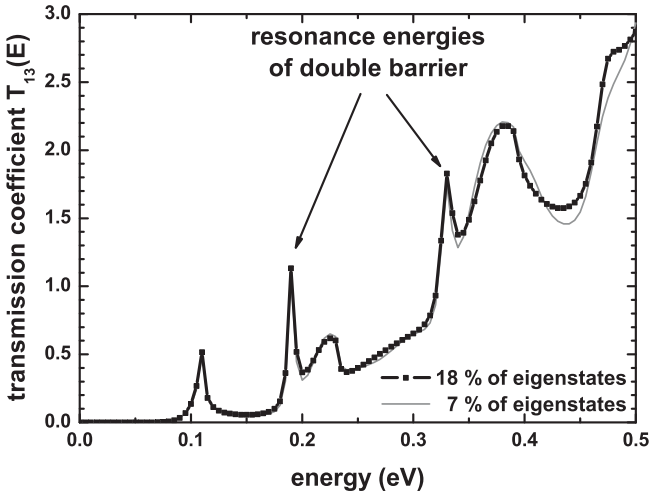


Figure 4.6.: Transmission coefficient $T_{13}(E)$ between lead 1 and lead 3 using an incomplete set of eigenstates of 7% (thin line) and 18% (thick solid line with squares) of the 2D device Hamiltonian of the closed system for the structure presented in Fig. 4.5. The first peak at 11 meV is a resonance due to electrons traveling the path below the Gaussian shaped barrier, the next two peaks are resonances where the electrons travel the other path where they tunnel through the double barrier.

transmission can certainly exceed $T = 1$ but the upper limit is $T = 41$. Figure 4.7 shows the calculated lead modes (eigenfunctions of the one-dimensional Schrödinger equation) of lead no. 1 of the same structure. The conduction band edge profile at the contact grid points (squares) is not constant due to the Gaussian shaped barrier in the center of the device that extends to the contacts. Shown are the lowest four eigenenergies (thin, constant lines) and their corresponding probability amplitudes $\left|\chi_m^{\lambda=1}\right\rangle^2$ that are shifted with their eigenenergies. The lead modes have been calculated by discretizing the 1D Schrödinger equation with a grid spacing of 0.5 nm and 41 grid points, using Dirichlet boundary conditions. The lead modes of lead no. 3 are identical because the structure is symmetric.

4.5.5. Transmission function of a nanowire structure (3D example)

Here we calculate the transmission of a three-dimensional example. We use a simple GaAs nanowire of cuboidal shape with a diameter of 10 nm × 10 nm and a length of 20 nm (see Fig. 4.8).

The device consists of two leads. They are represented by two-dimensional planes with dimensions 10 nm × 10 nm. Each lead has a total of 121 grid points (11 × 11 grid points). In each lead a two-dimensional Schrödinger equation has to be solved to obtain the eigenenergies and wave functions of the lead modes. The device region consists of

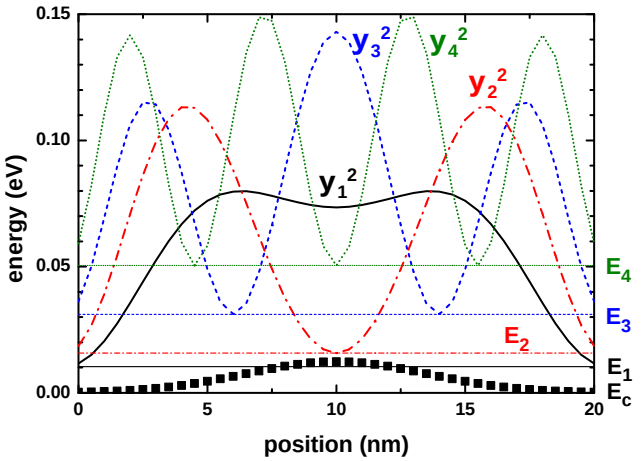


Figure 4.7.: Calculated lead modes (eigenfunctions of the one-dimensional Schrödinger equation) of lead no. 1 of the same structure as in Fig. 4.5. The conduction band edge profile E_c at the contact grid points (squares) is not constant due to the Gaussian shaped barrier in the center of the device. Shown are the lowest four eigenenergies (thin, constant lines) and their corresponding probability amplitudes that are shifted with their eigenenergies.

$11 \times 11 \times 21 = 2541$ grid points, which is equivalent to a grid spacing of 1.0 nm in all directions. Therefore the device Hamiltonian is a matrix of size 2541×2541. The conduction band edge profile is assumed to be constant and set to $E_c = 0\,\mathrm{eV}$. Also, the effective electron mass is constant throughout the device and equals $0.067 m_0$.

Figure 4.9 shows the calculated transmission coefficient $T(E)$ between the leads 1 and 2 as a function of energy. For the blue line 23.6% (600 of 2541) of all eigenvectors were

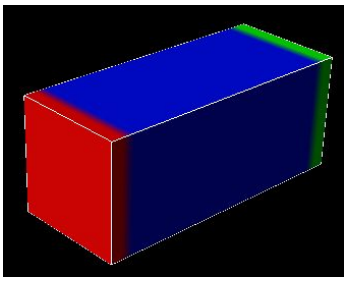


Figure 4.8.: Schematic sketch of the 3D nanowire showing the GaAs region (blue) that is placed between two contacts (red and green leads).

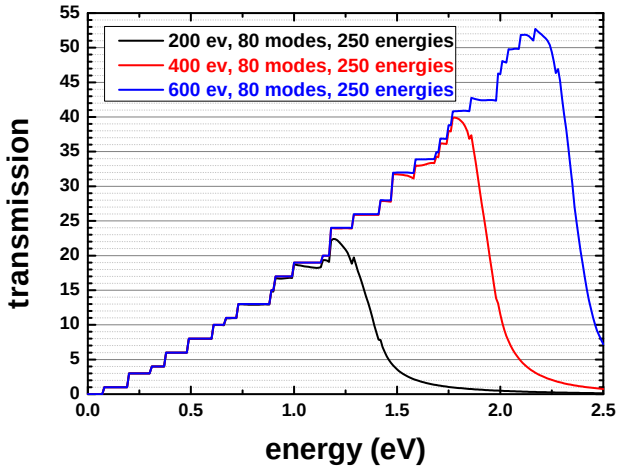


Figure 4.9.: Transmission coefficient $T(E)$ of a GaAs nanowire as a function of energy for three different percentages of eigenvectors taken into account.

used whereas for the red line only 15.7% (400 of 2541) had to be calculated. For the black line only 7.9% (200 of 2541) of all eigenvectors were used. Thus, a small percentage of eigenvalues suffices for the transmission function in the relevant energy range of interest, i.e. one does not have to calculate all eigenvectors of the device Hamiltonian which grossly reduces CPU time. Note that the transmission drops significantly once the cutoff energy of the highest eigenvector taken into account is reached.

The same data as in Fig. 4.9 is shown again in Fig. 4.10 as a zoom into the energy range 0 eV – 0.5 eV. The colored figures show the wave function amplitude of the lowest energy lead modes. Once the energy reaches 78 meV, the first lead mode energy is reached and then this mode transmits perfectly, giving a transmission of 1. The second and third lead mode states are degenerate due to the symmetry of the lead cross section, thus they have the same energy (191 meV). Consequently, once the energy of 191 meV is reached, the transmission increases by 2. The total transmission is now equal to 3 as all lead modes transmit perfectly. The energy of the 4^{th} lead mode is at 305 meV. The degeneracy of the 5^{th} and 6^{th} mode is accidental. They have the same energy. As one can clearly see, in this low energy limit, it is sufficient to calculate only a few percent of all eigenfunctions of the device Hamiltonian. For the leads, i.e. lead Hamiltonians, in all cases 80 of 121 (66.1%) eigenstates have been calculated.

Figure 4.11 shows the same transmission data as before but this time including the density of states. Again, the colors indicate taking into account 200, 400 or 600 eigenvectors of the decoupled system (closed system). The shape of the density of states is the typical one of two-dimensional confinement in quantum wires.

Technically, within the next**nano**^{3} software, for each contact (lead), a quantum clus-

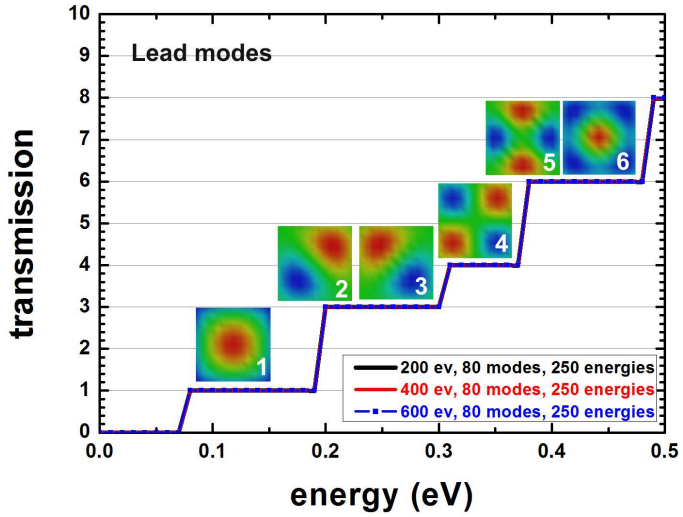


Figure 4.10.: Transmission coefficient $T(E)$ of a GaAs nanowire and relevant low energy lead modes. The energy of the lead modes is related to the steps in the transmission curve.

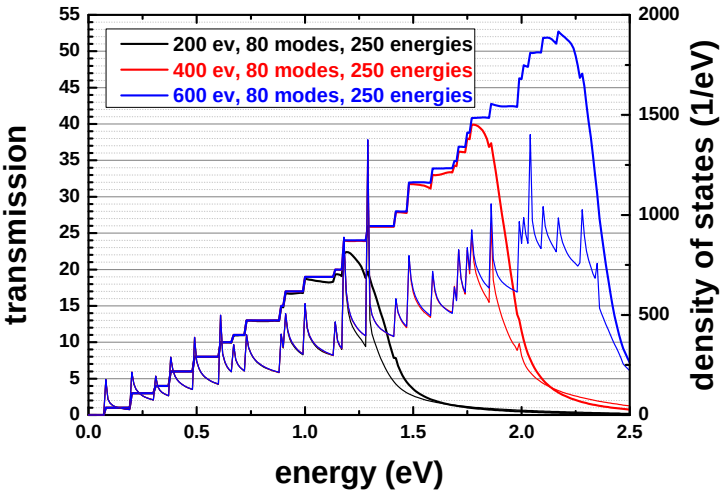


Figure 4.11.: Transmission coefficient $T(E)$ of a GaAs nanowire and corresponding density of states showing the typical shape of two-dimensional confinement.

ter ('lead quantum cluster') has to be defined because in each lead a two-dimensional Schrödinger equation has to be solved which gives us the lead modes (i.e. energies and

eigenvectors of the leads). In addition, a quantum cluster is required for the device itself ('main quantum cluster'). For each quantum cluster, the number of eigenstates to be calculated has to be specified.

The CBR algorithm is extremely well suited for parallelization, i.e. the calculation of $T(E)$ can be (ideally) parallelized for each energy step E. For each energy E (250 energy steps in this example) a matrix of size 160×160 has to be inverted. The size of 160 is determined by the sum over the number of lead modes taken into account for each lead. The upper limit would be the number of grid points in each lead that are in contact to the device, i.e. in this example where each lead has $11 \times 11 = 121$ grid points, the maximum size of the matrix to be inverted could be $242 = 121 + 121$. Here, only 80 eigenvalues for each lead were used.

I implemented an OpenMP parallelization which is the method of choice for multi-core processors. For 200, 400 and 600 eigenvalues the calculations took roughly 20, 25 and 34 seconds for the whole simulation including overhead (approx. 5 seconds for processing input file, generating output files, ...) on a 12-core computer. On a single-core CPU, the respective times were 44, 58 and 77 seconds. The eigenvectors of the device Hamiltonian have been calculated with the ARPACK eigenvalue solver [LSY98]. We note that in this example the time-consuming part was the calculation of the eigenvectors, which itself is not parallelized[2] with the exception of calls to BLAS [BLA] and LAPACK [LAP] routines, and not the implementation of the (ideal) OpenMP-parallelization of the $T(E)$ calculation.

As an alternative to OpenMP, I also implemented a Coarray Fortran parallelization which allows us to connect several multi-core CPUs over a network using the g95 Fortran compiler[3] which showed similar performance as the OpenMP implementation.

4.6. Local density of states

In order to obtain the local density of states (LDOS) for each energy E_i, some additional steps are required. In the following the subscript C indicates that the matrices have size N_C. If an incomplete set of lead modes has been used (see Subsection 4.5.2) all these matrices are in fact smaller and have only a size of $N_\mathrm{C,m}$. However, for better readability we now omit the subscript m and write only C.

We need the lower left part $\mathbf{G}^\mathrm{R}_\mathrm{DC}$ of the retarded Green's function that correlates the device and the contacts (see eq. (4.6)). It is obtained from the corresponding Dyson equation

$$\mathbf{G}^\mathrm{R}_\mathrm{DC} = \mathbf{G}^0_\mathrm{DC}\mathbf{B}^{-1}_\mathrm{C} \tag{4.42}$$

$$\mathbf{B}_\mathrm{C} = \mathbf{1}_\mathrm{C} - \mathbf{\Sigma}_\mathrm{C}\mathbf{G}^0_\mathrm{C}, \tag{4.43}$$

where $\mathbf{1}_\mathrm{C}$ is the identity matrix and $\mathbf{\Sigma}_\mathrm{C}\mathbf{G}^0_\mathrm{C}$ is a simple matrix multiplication of two small matrices that have been calculated already (see Subsection 4.4.3 and Subsection 4.4.3).

[2] A parallel version of ARPACK (Parallel ARPACK, PARPACK) [MS96] which is suitable for distributed memory parallel architectures is also available.
http://www.caam.rice.edu/~kristyn/parpack_home.html

[3] http://www.g95.org

The matrices $\mathbf{G}^{\mathrm{R}}_{\mathrm{DC}}$ and $\mathbf{G}^{0}_{\mathrm{DC}}$ are not square matrices. As they correlate the interior device grid points with the leads, they have the dimension $N_{\mathrm{D}} \times N_{\mathrm{C}}$. They are represented within a mixed real space and lead mode space representation. The Green's function $\mathbf{G}^{0}_{\mathrm{DC}}$ of the closed device can be expressed in terms of the following spectral representation

$$\mathbf{G}^{0}_{\mathrm{DC}} = \sum_{n=1}^{n_\alpha} \frac{|\psi_{n,\mathrm{D}}\rangle \langle \psi_{n,\mathrm{C}}|}{E_i - E_n}, \tag{4.44}$$

which reads in mixed real space (index z) and lead mode space (index m)

$$\mathbf{G}^{0}_{\mathrm{DC}}(z,m) = \sum_{n=1}^{n_\alpha} \frac{\langle z|\psi_{n,\mathrm{D}}\rangle \langle \psi_{n,\mathrm{C}}|m\rangle}{E_i - E_n} \tag{4.45}$$

$$= \sum_{n=1}^{n_\alpha} \frac{\psi_{n,\mathrm{D}}\ \chi_n^{\mathrm{T}}}{E_i - E_n} = \mathbf{\Psi}_{\mathrm{D}}\mathbf{M}. \tag{4.46}$$

In fact, this only involves a matrix multiplication $\mathbf{\Psi}_{\mathrm{D}}\mathbf{M}$ where the dimensions of the matrices are

$$\mathbf{G}^{0}_{\mathrm{DC}}: \quad N_{\mathrm{D}} \times N_{\mathrm{C}} \tag{4.47}$$

$$\mathbf{\Psi}_{\mathrm{D}}: \quad N_{\mathrm{D}} \times n_\alpha \tag{4.48}$$

$$\mathbf{M}: \quad n_\alpha \times N_{\mathrm{C}}. \tag{4.49}$$

The matrix $\mathbf{\Psi}_{\mathrm{D}}$ contains in columns $1, ..., n_\alpha$ the wave vectors $\psi_{n,\mathrm{D}}$ of the eigenstate n of the closed device Hamiltonian

$$\mathbf{\Psi}_{\mathrm{D}} = \begin{pmatrix} \psi_1 & ... & \psi_{n_\alpha} \end{pmatrix}. \tag{4.50}$$

The matrix $\mathbf{M}$ is defined as

$$\mathbf{M} = \begin{pmatrix} M_1^{\mathrm{T}} \\ ... \\ M_{n_\alpha}^{\mathrm{T}} \end{pmatrix}, \tag{4.51}$$

where we store for each eigenvalue E_n the following vector

$$M_n = \frac{1}{E_i - E_n}\chi_n. \tag{4.52}$$

This is actually the same as eq. (4.20) (or eq. (4.35), respectively) apart from the coefficient $1/(E_i - E_n)$. $\mathbf{M}$ is of dimension $n_\alpha \times N_{\mathrm{C}}$ where $n_\alpha \leq n_{\mathrm{T}}$ is the number of eigenvalues taken into account, and N_{C} is the dimension of the χ_n vectors in the mode space representation.

The inverted matrix $\mathbf{B}_{\mathrm{C}}^{-1}$ can be obtained using the same implementation of an inversion algorithm analogously to Subsection 4.4.3. The matrix multiplications involving large matrices (eq. (4.42), eq. (4.46)) can be performed efficiently using standard routines from numerical libraries (e.g. BLAS routines ZGEMM and DGEMM [BLA], respectively).

The local density of states $\rho(z, E)$ is the diagonal (divided by 2π) of a more general concept called the spectral function $A(E) = A(z, z', E)$. The local density of states ρ^λ for each lead λ (lead connected local density of states) can easily be calculated at each grid point z from the retarded Green's function $\mathbf{G}^{\mathrm{R}}_{\mathrm{DC}}(z, m)$ (in mixed real space and mode space representation) and the broadening matrix $\mathbf{\Gamma}$

$$\rho^\lambda(z, E_i) = \frac{1}{2\pi} A_\lambda(z, z, E_i) \tag{4.53}$$

$$= \frac{1}{2\pi} \left\langle z \left| \mathbf{G}^{\mathrm{R}} \mathbf{\Gamma}^\lambda \mathbf{G}^{\mathrm{R}\dagger} \right| z \right\rangle \tag{4.54}$$

$$= \frac{1}{2\pi} \sum_{m=1}^{m^\lambda_\alpha} \left| \mathbf{G}^{\mathrm{R}}_{\mathrm{DC}}(z, m) \right|^2 \mathbf{\Gamma}^\lambda_{mm}. \tag{4.55}$$

$\mathbf{\Gamma}^\lambda$ is diagonal in the subspace of the propagating lead modes. The sum runs over all modes m of lead λ. For a 1D simulation, there is only one mode, and thus $\mathbf{G}^{\mathrm{R}}_{\mathrm{DC}}$ is a $N_{\mathrm{D}} \times 2$ matrix and $\mathbf{\Gamma}^\lambda_{mm}$ is the diagonal of $\mathbf{\Gamma}^\lambda$, i.e. only a vector of length 2 has to be stored. The LDOS must have units of 1/energy 1/length ($\mathrm{eV}^{-1}\mathrm{nm}^{-1}$) in a 1D simulation (2D: 1/energy 1/area, 3D: 1/energy 1/volume). As we normalized the wave functions to be dimensionless (eq. (4.16), eq. (4.17)), the calculated LDOS has to be divided by the grid spacing Δ for consistency. The total local density of states is simply the sum over the LDOS of each lead

$$\rho(z, E_i) = \sum_{\lambda=1}^{L} \rho^\lambda(z, E_i). \tag{4.56}$$

So far we calculated the local density of states only at the interior device grid points N_{D}. All of the equations in this section apply equally well to the contact grid points N_{C} if one replaces the subscript D with the subscript C. In a numerical implementation, one simply has to use in eq. (4.46) the wave vectors $\psi_{n,\mathrm{T}}$ of the total device

$$\mathbf{G}^0_{\mathrm{TC}}(z, m) = \mathbf{\Psi}_{\mathrm{T}} \mathbf{M}, \tag{4.57}$$

and to replace in eq. (4.42) and eq. (4.55) $\mathbf{G}^{\mathrm{R}}_{\mathrm{DC}}$ by $\mathbf{G}^{\mathrm{R}}_{\mathrm{TC}}$ to obtain the local density of states $\rho^\lambda(z, E_i)$ for both interior and contact grid points simultaneously. The matrix $\mathbf{\Psi}_{\mathrm{T}}$ is stored in memory and has been obtained from numerically solving the Schrödinger equation (eq. (4.14)).

Density of states The density of states (DOS) $D^\lambda(E_i)$ for each lead can be obtained by integrating the local density of states $\rho^\lambda(z, E_i)$ for each energy E_i over the spatial coordinate z

$$D^\lambda(E_i) = \int \rho^\lambda(z, E_i)\, \mathrm{d}z = \Delta \sum_{z=1}^{N_{\mathrm{T}}} \rho^\lambda_z(E_i). \tag{4.58}$$

Thus the DOS $D^\lambda(E_i)$ can easily be obtained by adding the components of the vector that stores $\rho^\lambda_z(E_i)$ and multiplying this sum by the grid spacing Δ. The total density of

states is then simply the sum over the DOS for each lead

$$D(E_i) = \sum_{\lambda=1}^{L} D^{\lambda}(E_i). \tag{4.59}$$

The DOS is in units 1/energy.

4.7. Density

The charge density can be calculated via the density matrix or via the local density of states [VMK+08]. We recommend to use the local density of states. This is favorable within a self-consistent scheme, since it allows for the use of a predictor–corrector scheme (see Subsection 4.8.2) to improve the convergence. From the lead connected local density of states $\rho^{\lambda}(\mathbf{x}, E)$, the local energy resolved carrier density $n_E^{\lambda}(\mathbf{x}, E)$ for each lead λ is obtained by occupying each level with the distribution function $f(E, \mu_\lambda)$ of the corresponding lead

$$n_E^{\lambda}(\mathbf{x}, E) = g_{\mathrm{s}} g_{\mathrm{v}} \rho^{\lambda}(\mathbf{x}, E) f_{d\mathrm{D}}\left(E, \mu_\lambda\right), \tag{4.60}$$

where $g_{\mathrm{s}} = 2$ is the spin degeneracy and g_{v} is the valley degeneracy. The latter is relevant when treating electrons that are in the X or L valleys, like in AlAs, silicon or germanium. In higher dimensions or if these bands are split due to strain, usually for each valley a separate Schrödinger equation has to be solved. Depending on the simulation dimension ($d = 1, 2, 3$) the appropriate Fermi function $f_{d\mathrm{D}}$ has to be used which takes into account the $\mathbf{k}_{\parallel}$ vectors that occur in 1D and 2D simulations. For a device that is homogeneous along the x and y directions (1D simulation) it is given at a particular energy E_z by

$$f_{1\mathrm{D}}(E_z, \mu) = \frac{m_{\parallel} k_{\mathrm{B}} T}{2\pi\hbar^2} \ln\left(1 + e^{-(E_z - \mu)/k_{\mathrm{B}}T}\right), \tag{4.61}$$

where $m_{\parallel}(z)$ is the effective mass tensor component in the (x, y) plane of the respective valley (which generally varies with position z and thus has to be averaged over the spatial coordinates weighted with the local density of states for each energy). $f_{1\mathrm{D}}$ is in units of 1/area.

The analogous equations for 2D and 3D devices are

$$f_{2\mathrm{D}}(E_{x,y}, \mu) = \sqrt{\frac{m_{\parallel} k_{\mathrm{B}} T}{2\pi\hbar^2}} F_{-1/2}\left((\mu - E_{x,y})/k_{\mathrm{B}}T\right), \tag{4.62}$$

where $m_{\parallel}(x, y)$ is the effective mass tensor component along the homogeneous z direction, and $F_{-1/2}$ is the Fermi–Dirac integral of order $-1/2$ which can be evaluated efficiently using approximation formulas [Ant93]. $f_{2\mathrm{D}}$ is in units of 1/length. In 3D the usual Fermi function is used which is of course dimensionless

$$f_{3\mathrm{D}}(E, \mu) = \frac{1}{1 + \exp\left((E - \mu)/k_{\mathrm{B}}T\right)}. \tag{4.63}$$

From the lead connected local energy resolved density $n_E^\lambda(\mathbf{x}, E)$, the local carrier density $n^\lambda(\mathbf{x})$ for each lead λ is obtained by integrating over the energy E

$$n^\lambda(\mathbf{x}) = \int n_E^\lambda(\mathbf{x}, E)\ \mathrm{d}E. \tag{4.64}$$

The total density is the sum over the contributions from all leads

$$n(\mathbf{x}) = \sum_{\lambda=1}^{L} n^\lambda(\mathbf{x}). \tag{4.65}$$

The units are 1/volume in all dimensions.

In the explanations above we introduced the term *energy resolved* density. For 1D and 2D simulations this energy $E = E_z$ did not take into account the energy due to $\mathbf{k}_\parallel \neq 0$. The total energy of the electron is given by eq. (4.21) for the 1D case and by eq. (4.37) for the 2D case. It is necessary to include these $\mathbf{k}_\parallel$ contributions into the energy resolved density to get meaningful plots. This is done by first evaluating the local density of states $\rho(z, E_{\text{total}})$ for the total energy, and then occupying the LDOS by the usual Fermi function (eq. (4.63)). In 1D simulations, information about the system under study can be obtained by plotting the energy resolved electron density $n(\mathbf{x}, E_{\text{total}})$ and the energy resolved electron density $n^\lambda(\mathbf{x}, E_{\text{total}})$ for each lead. These are two-dimensional plots like the local density of states. The density can be split into two parts, one originating from the left lead, and one from the right lead (see Fig. 4.14). In 2D simulations the plot of the energy resolved electron density or local density of states is a three-dimensional plot. This makes it difficult to analyze these quantities in 3D simulations where they are four-dimensional. Thus one can only plot slices through these 4D data.

4.8. Self-consistent CBR algorithm

The self-consistent solution of the ballistic transport properties of an open device requires the repeated solution of the Schrödinger and Poisson equations due to the coupling via the potential and the quantum mechanical density. Also, the lead modes are calculated self-consistently using the potential at the contacts, obtained from the solution of the Poisson equation. In principle, it is possible to simply iterate the solution of theses equations, and with enough damping this will lead to yield a converged result. To improve the convergence of a highly nonlinear set of coupled equations, such as the Schrödinger–Poisson problem, the Newton algorithm is usually the first choice. But since the exact Jacobian cannot be derived analytically and a numerical evaluation would be too costly, the simple adaption of this method is not feasible. For the case of a closed system this problem has been solved using a predictor–corrector approach [TGPR97]. The aim of this highly efficient method is to find a good approximation for the quantum density as a function of the electrostatic potential where an expression for the Jacobian is known. Within this approximation the nonlinear Poisson equation can efficiently be solved using the Newton scheme resulting in a predictor update for the electrostatic

potential and the carrier density. If this approximation is close enough to the real quantum density, only very few updates will be necessary to yield a converged solution. This means that for each bias step, the Schrödinger equation has to be solved less than approximately 10 times until the potential and the density are sufficiently converged. In the *nin*–resistor example presented below, the Schrödinger equation had to be solved only 2–3 times for each bias step.

4.8.1. Poisson equation in the CBR method

The Poisson equation (eq. (1.1)) describes the electrostatics within the device and has been described in Section 1.1. In all CBR calculations in this thesis, we only take into account one conduction band and calculate the electron density n quantum mechanically as described in Section 4.7. Here, we only consider fully ionized donors N_{D}^{+} (Section 1.2) and neglect all other contributions to the density. For the CBR calculations we discretize the Poisson equation on a uniform grid. For both equilibrium and nonequilibrium calculations, we use Neumann boundary conditions for the Poisson equation which implies a vanishing electric field at the boundaries (eq. (1.3)). This is the recommendation for ballistic devices [Dat05]. An alternative would be to use Dirichlet boundary conditions for nonequilibrium simulations [LKBJ97]. Here, one first has to determine the electrostatic potential in equilibrium (built-in potential) using zero-field (Neumann) boundary conditions. The electrostatic potential at the boundaries is then fixed with respect to the chemical potentials taking into account the previously calculated built-in potential at the boundaries. For both boundary conditions, the chemical potentials at the contacts are fixed and correspond to the applied bias. Further boundary conditions are summarized in Ref. [LKF04]. These include the concept of a drifted Fermi distribution function in the leads that accounts for a net current flow in those leads.

4.8.2. Predictor–corrector approach

A fast and robust iterative method for obtaining self-consistent solutions to the coupled system of Schrödinger and Poisson equations is very important. Basically, a simple expression describing the dependence of the quantum electron density on the electrostatic potential is required ($\frac{\partial \rho}{\partial \phi}$). This expression is then used to implement an iteration scheme, based on a predictor–corrector type approach, for the solution of the coupled system of differential equations. Within the CBR method, a predictor–corrector approach can easily be applied making use of the previously calculated local density of states by modifying eq. (4.60) slightly to get the energy resolved density for the predictor potential. This predictor density $n^{\lambda}_{E,\mathrm{p}}(\mathbf{x}, E, \Delta_{\phi})$ is then given by

$$n^{\lambda}_{E,\mathrm{p}} = g_{\mathrm{s}} g_{\mathrm{v}} \rho^{\lambda}(\mathbf{x}, E) f_{d\mathrm{D}}\left(E - e\Delta_{\phi}(\mathbf{x}), \mu_{\lambda}\right). \tag{4.66}$$

The idea behind this approximation is that to first order the wave functions, and therefore the local density of states $\rho^{\lambda}(\mathbf{x}, E)$, remain unchanged for small deviations in the potential. Only the eigenenergies are adjusted locally to small changes in the electro-

static potential $\Delta_\phi(\mathbf{x})$. This is achieved by using $E - e\Delta_\phi(\mathbf{x})$ instead of E as the new argument for the Fermi function $f_{d\mathrm{D}}$.

The charge density used in the Poisson equation is a function of the electrostatic potential ('nonlinear' equation). The nonlinear Poisson equation can be solved very fast using a predictor density. This density avoids the time-consuming procedure of solving the Schrödinger equation many times. Once the new electrostatic potential for the predictor density has been obtained, the new quantum mechanical density, i.e. the new local density of states for this potential can be evaluated. This procedure is iterated until convergence of both the electrostatic potential and the quantum mechanical charge density is achieved.

The nonlinear Poisson equation itself is solved by a Newton–Raphson method where the functional

$$\mathbf{F} = \mathbf{A} \cdot \phi + \rho = 0 \tag{4.67}$$

is minimized. Here, $\mathbf{A}$ represents the discretized Poisson matrix and ρ is a vector representing the charge density for each grid point. The Newton algorithm finds an electrostatic potential vector $\phi^{j+1} = \phi^j + \Delta_\phi$ such that the magnitude of the residuum vector $\mathbf{F}$ becomes smaller than a certain small threshold of ε. The electrostatic potential ϕ^j of the j^{th} iteration step is kept fixed within the Newton method. The index j refers to the outer Schrödinger–Poisson iteration and counts how often the Schrödinger equation has to be solved until convergence is obtained. Once the Newton algorithm has converged to a correction Δ_ϕ, the Schrödinger equation (i.e. the CBR algorithm) is solved for the updated electrostatic potential ϕ^{j+1}. The new local density of states is then input to the next iteration of the Newton algorithm.

For the Newton correction, the Jacobi matrix $\mathbf{J}$ is needed. It is simply the Poisson matrix plus the derivative of the density with respect to the potential

$$\mathbf{J} = \frac{\partial \mathbf{F}}{\partial \phi} = \mathbf{A} + \frac{\partial \rho}{\partial \phi} = \mathbf{A} + \frac{\partial n_{\mathrm{p}}}{\partial \phi}. \tag{4.68}$$

Thus within the CBR method, the derivative of the predictor density n_{p} with respect to the potential is needed. This derivative is available using eq. (4.66) and the derivative of the Fermi functions $f_{d\mathrm{D}}$ with respect to Δ_ϕ.

The iteration approach presented in this section simplifies the numerical implementation of the nonlinear Schrödinger–Poisson problem significantly. In addition, it provides excellent convergence speed and stability. Further details about it can be found in Ref. [VMK$^+$08].

4.8.3. Self-adapting energy grid

For the numerical implementation of a self-consistent scheme using a continuous density of states, the energy grid is of high importance. To integrate the carrier density, we discretize the local density of states in energy space and then employ a simple numerical integration by summing up the values for each energy step weighted by the Fermi distribution and the energy grid spacing Δ_E. Since the DOS is a very spiky function

with peaks corresponding to highly localized states due to the onset of the conduction band edges at the contacts (1D) or due to the propagating lead mode energies (2D, 3D), it is very important for the convergence of the self-consistent CBR algorithm to have these features properly resolved. Additional peaks arise from quasibound states, like for instance in the double barrier structure (10 nm barrier widths) as discussed in Subsection 4.4.4. Usually the main structural features in the DOS are due to the lead modes. If quasibound states are the dominant features in the DOS, one could use the information about their energy levels (which is available within the CBR method) to optimize the energy grid. Thus we need an energy grid that is self-adapting to the density of states which varies for each iteration. Otherwise, a well converged self-consistent solution is not possible unless a lot of energy grid points are used.

In 2D simulations of e.g. a double gate MOSFET where the channel acts as a one-dimensional wire, the peaks show a $1/\sqrt{(E_i - E_m^\lambda)}$ dependence, where E_m^λ is the peak energy arising due to the onset of the lead modes. In 2D and 3D there are always several onsets as there are typically many lead modes involved. The peaks in our 1D *nin*–resistor example (Fig. 4.15) show also a $1/\sqrt{(E_i - E_m^\lambda)}$ dependence, where $E_m^\lambda = E_c^\lambda$ is the peak energy arising due to the onset of the conduction band edges at the contacts. In 1D simulations there are only two onsets, one for each contact. The integral over the peak is thus very poor when using an energy grid with constant grid spacing (uniform energy grid, see Subsection 4.4.3), since the relative distance between the nearest energy grid point E_i and the peak energy E_m^λ is arbitrary. Additionally, the lead mode energy is slightly shifted with each iteration step, leading to a varying integration error during the self-consistent cycle, which is an extreme handicap for any self-consistent algorithm. Thus a solution to this problem is to use the physical information we have about the system and employ a self-adapting energy grid that resolves each known (i.e. relevant) peak m with a local energy grid of a few tens of energy grid points that is fixed to the lead mode energy E_m^λ. Additionally, extra points are distributed in the space between the peaks to obtain a smooth enough energy grid. An exponential grid type is recommended since it provides a good resolution of the $1/\sqrt{E}$ behavior of the peaks. In order to avoid singularities the energy grid points are not allowed to match exactly the eigenenergies of the closed system. For each peak, the first grid point is set slightly below the onset of the peak and then each grid point i is set with increasing energy grid spacing

$$\Delta_{E^i} = g\ \Delta_{E^{i-1}} = g^i\ \Delta_{E^0}, \tag{4.69}$$

starting with the initial grid spacing $\Delta_{E^0} = 0.1\,\mathrm{meV}$, and a grid factor $g = 1.2$, for instance. A grid factor of $g = 1.0$ leads to a locally linear grid which has been found to be not as efficient as the exponential grid. The parameters that specify the energy grid are the total number of energy grid points, the maximum number of peaks taken into account, the number of energy grid points in the local grid around a peak, and the grid factor. The minimum value of the energy grid should be slightly below the minimum of the conduction band edges of the contacts, the maximum value should not be higher than $E_{\max} = 0.25t$, where t is defined analogously to eq. (4.24).

Figure 4.15 demonstrates that the peaks in the LDOS and DOS of a simple *nin*–structure (see Subsection 4.8.5) are well resolved, and that for other regions in the energy

interval less grid points are fully sufficient. The energy grid consists of 300 grid points including the extra points used to resolve the onsets of the two peaks at the conduction band edges of the contacts. Importantly, the integration error is reduced compared to the uniform grid and remains constant within the iteration, since the grid is locally fixed to the shifting lead mode energies (or conduction band edges in a 1D simulation). The convergence behavior of a uniform grid with an order of magnitude more energy grid points is very similar for the first iteration steps. The achieved convergence is measured by a residuum which is a very small number. Compared to the self-adapting grid, the uniform grid reaches a bottom at the residuum, which cannot be reduced further. This is due to the fluctuating integration error. In contrast, the self-adapting energy grid guarantees satisfying convergence.

As the contact block matrices have to be inverted for each energy, the computational time depends linearly on the total number of energy grid points. Therefore, a numerical implementation of an optimized energy grid is very important for an efficient use of the CBR method.

4.8.4. Extracting the quasi-Fermi level

For all calculations presented in this paper, the extraction of the quasi-Fermi level was not necessary because only one conduction band has been involved. For equilibrium solutions, we so far assumed that the Fermi level (chemical potential) is constant and fixed to $E_\mathrm{F} = 0\,\mathrm{eV}$, allowing the semiconductor band edges to adjust according to the electrostatic potential as calculated from the Poisson equation (see Subsection 4.8.1). For nonequilibrium calculations where the device is under bias, one could extract a spatially varying quasi-Fermi level $E_\mathrm{F}(z)$ in order to get meaningful (or to avoid artificially wrong) charge densities for all other bands that are not treated quantum mechanically with the CBR method, e.g. hole bands or higher lying electron bands. This might be necessary for the self-consistent CBR algorithm under high bias conditions, where for each iteration the quasi-Fermi levels have to be obtained self-consistently. The reason is that the equation for the classical densities needs a reasonable value for the local quasi-Fermi level. In nonequilibrium calculations electrons and holes can be described by different quasi-Fermi levels ($E_{\mathrm{F,n}}(z)$, $E_{\mathrm{F,p}}(z)$, respectively). The quasi-Fermi level for electrons can be obtained by finding (e.g. using a bisection algorithm) for each grid point z the appropriate local quasi-Fermi level $E_{\mathrm{F,n}}(z)$ that corresponds to the actual electron density at this grid point (and similar for the holes). These Fermi levels would lie in between the chemical potentials of the left and right contact which are kept fixed in a nonequilibrium calculation. Rather than occupying the lead connected local density of states with the chemical potential of the relevant lead (eq. (4.60)), one would occupy the total local density of states (eq. (4.56)) at position z by taking a suitable average ($E_{\mathrm{F,n}}(z)$) of the chemical potentials of all leads.

Bound states treatment Electronic states that are below the conduction band edges of the contacts do not get occupied within a ballistic algorithm. All higher lying states contribute via the local density of states to the quantum mechanical density. It is not a re-

alistic treatment to ignore the lower lying bound states as they usually get filled through scattering events. Therefore the density originating from the bound states obviously contributes to the electrostatics of the device and should be included into the Poisson equation. An example calculation of a quantum well that is completely empty within a ballistic calculation but gets filled once scattering is included has been discussed in detail in Ref. [KTV05]. As the probability densities of the electronic states are available (see Subsection 4.4.1), one could use this information and occupy the states that are below the conduction band edges of the contacts locally with a self-consistently determined quasi-Fermi level. This is the standard approach usually employed in Schrödinger–Poisson solvers. Here, however, for energies where the LDOS from the ballistic calculation is available, the CBR density is used instead. So the total density has two contributions, one from the bound states and one from the CBR density. Another approach how to include bound states is described in Ref. [VMK$^+$08].

4.8.5. *nin*–resistor (1D example)

As a simple example to illustrate the self-consistent CBR method, we choose a *nin*–structure where quantum confinement effects are not relevant. Hence, the equilibrium solution can easily be checked against the standard approach for calculating the carrier concentration in semiconductor devices. This classical density is obtained using Fermi–Dirac integrals and the effective density of states of the conduction (and valence) bands. We emphasize that in Fig. 4.12 and Fig. 4.13 we only used quantum mechanical densities calculated with the CBR method (see Section 4.7). In equilibrium, the CBR approach leads to the same conduction band profile and the same carrier densities as the classical approach (not shown). The *nin*–structure consists of GaAs and has a length of 80 nm. The doping profile is symmetric with a donor concentration of $N_\mathrm{D}^+ = 1 \cdot 10^{18}$ cm^{-3} (fully ionized). The 35 nm wide n–type doped regions at the source and drain sides are separated by a 10 nm wide intrinsic region in the middle. For comparison, we also study an asymmetrically doped n^+in–structure where the left n–type region has a doping concentration of $N_\mathrm{D}^+ = 2 \cdot 10^{18}$ cm^{-3} (0-35 nm) and the right doping region has a concentration of $N_\mathrm{D}^+ = 1 \cdot 10^{18}$ cm^{-3} (45-80 nm). The temperature is set to 300 K. The device is discretized with a grid spacing of 1 nm. A self-adapting energy grid of 300 energy grid points is used. For each bias point, it is sufficient to solve the Schrödinger equation only 2–3 times to get well converged results. This shows that this *nin*–resistor is well suited as a test case to benchmark an efficient implementation of both the self-adapting energy grid, and the predictor–corrector algorithm.

Figure 4.12 and Fig. 4.13 show the conduction band edge profiles and electron densities of the symmetric *nin*– and asymmetric n^+in–structures, respectively, calculated with the self-consistent CBR method. The solid lines are equilibrium results and the dotted lines correspond to an applied bias of $V_\mathrm{SD} = -50$ mV at the right contact. The chemical potential in equilibrium is equal to $\mu = 0$ meV (dashed line). Under bias, the chemical potential of the right contact is increased by 50 meV, indicated by the vertical arrows. As a consequence of the zero-field boundary conditions for the Poisson equation, the band edges are flat at the contacts. However, for the symmetric *nin*–structure the

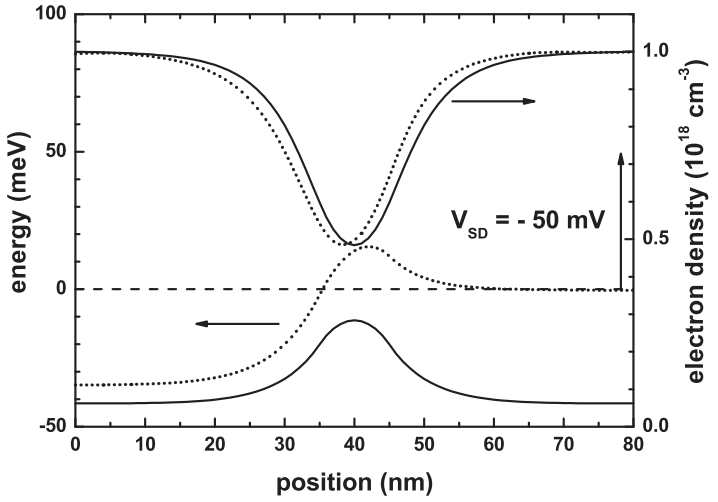


Figure 4.12.: Conduction band edge profiles and electron densities of a symmetric *nin*–structure calculated with the self-consistent CBR method. Solid lines are equilibrium results, dotted lines correspond to an applied bias of $V_{\mathrm{SD}} = -50\,\mathrm{mV}$ at the right contact. The chemical potential in equilibrium is equal to $\mu = 0\,\mathrm{meV}$ (dashed line). Under bias, the chemical potential of the right contact is increased by 50 meV, indicated by the vertical arrow. The doping profile is symmetric ($N_{\mathrm{D}}^{+} = 1 \cdot 10^{18}$ cm^{-3}).

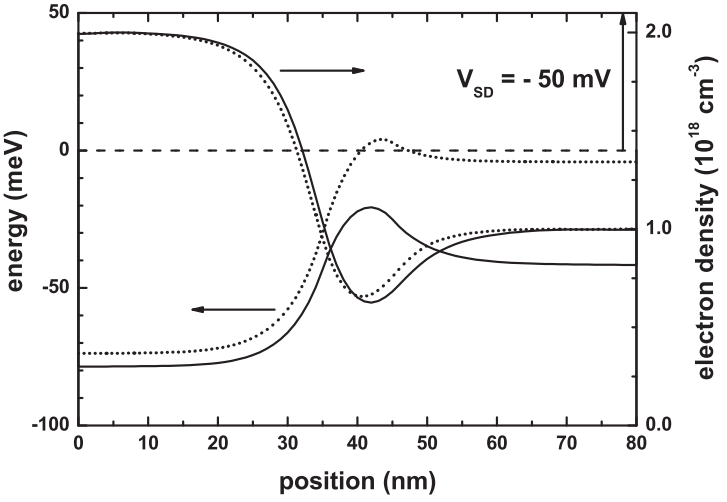


Figure 4.13.: Same as Fig. 4.12 but for the $n^{+}in$–structure that has an asymmetric doping profile (0-35 nm: $N_{\mathrm{D}}^{+} = 2 \cdot 10^{18}$ cm^{-3}, 45-80 nm: $N_{\mathrm{D}}^{+} = 1 \cdot 10^{18}$ cm^{-3}).

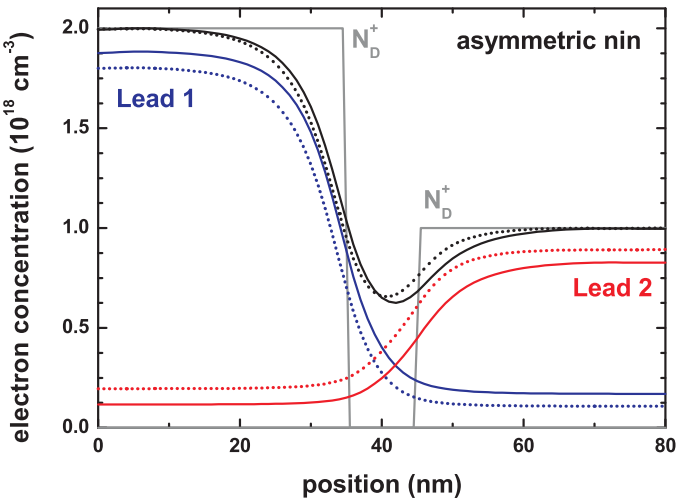


Figure 4.14.: Individual electron densities due to the left (blue lines) and the right (red lines) contact are shown for the asymmetric n^+in–structure. The total density is identical to Fig. 4.13 (black lines). Solid lines are equilibrium results, dotted lines correspond to the applied bias $V_{\mathrm{SD}} = -50\,\mathrm{mV}$. The donor concentration profile N_{D}^+ is also shown. Raising the chemical potential at the right contact increases (decreases) the density due to the occupation of the corresponding local density of states of the right (left) contact.

difference in the conduction band edges at the left and right contact is smaller than the actual difference in the chemical potentials. The same holds for the asymmetric n^+in–structure if one takes the built-in potential (of the equilibrium calculation) into account. The reason for this behavior is as follows (see Section 11.4 “Where is the voltage drop” of Ref. [Dat05]): In ballistic simulations a fraction of the density of states at one contact is always controlled by the contact at the other end. Making the end regions of the device longer will not change this situation. This can easily be understood by separately visualizing the electron densities that originate from the left and right contacts. This is shown in Fig. 4.14 where the individual electron densities due to the left (blue lines) and the right (red lines) lead are shown for the asymmetric n^+in–structure. The total density is identical to Fig. 4.13 (black lines). Solid lines are equilibrium results, dotted lines correspond to the applied bias $V_{\mathrm{SD}} = -50\,\mathrm{mV}$. The donor concentration profile N_{D}^+ is also shown. Raising the chemical potential at the right contact increases the density due to the occupation of the corresponding local density of states of this contact (lead connected local density of states, see Section 4.6). Consequently, the density due to the other lead must decrease to guarantee global charge neutrality. There are two ways for the density to decrease, one is changing the chemical potential of the relevant lead (which is not possible as it is fixed due to the boundary condition), the other possibility is to adjust the electrostatic potential, and thus the conduction band edge. The latter

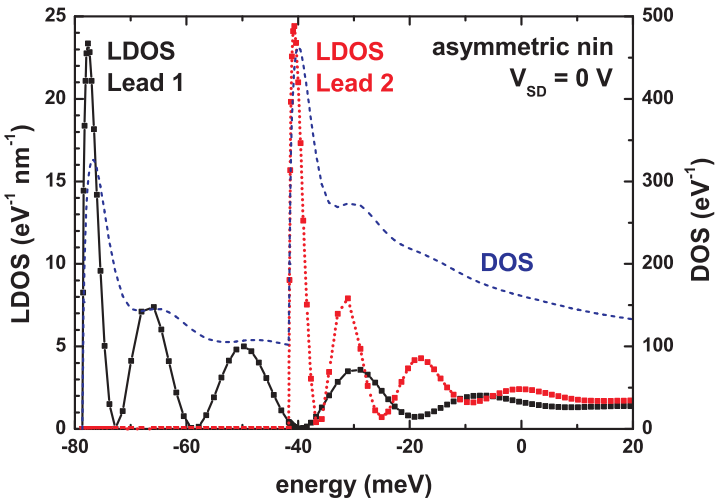


Figure 4.15.: Local density of states (LDOS) at the left (Lead 1, black solid line) and right (Lead 2, red dotted line) contact of the asymmetric n^+in–structure in equilibrium. The self-adapting energy grid is able to resolve the peaks in the LDOS sufficiently accurate (300 energy grid points in total). Also shown is the density of states (DOS, blue dashed line) which is the integrated LDOS over the position (sum over all lead contributions). The DOS has peaks at the onset of the conduction bands edges at the left and right contacts.

situation corresponds to zero-field boundary conditions (Neumann). This explains why Dirichlet boundary conditions are inappropriate for ballistic devices. For quantum cascade laser (QCL) simulations where the doping concentration is low, Neumann boundary conditions seem to be a natural choice where one allows the derivative of the potential at the left and right boundaries

$$\frac{\partial \phi}{\partial z} = \text{const} \tag{4.70}$$

to adjust self-consistently under the condition of global charge neutrality, i.e. requiring equal slope at the boundaries. The slope is adjusted in such a way that the potential drop across the device equals the bias voltage that is defined by the difference between the chemical potentials in the contacts [KYV+09]. This will lead to finite electric fields at the boundaries that correspond to the applied electric field in the QCL.

Figure 4.15 shows the local density of states at the left (Lead 1, black solid line) and right (Lead 2, red dotted line) contact of the asymmetric n^+in–structure in equilibrium, i.e. one-dimensional slices at the first ($z = 1$) and last ($z = N_\mathrm{T}$) grid point of the two-dimensional LDOS $\rho^\lambda(z, E_i)$ plot of lead λ (eq. (4.55)). The self-adapting energy grid is able to resolve the peaks in the LDOS sufficiently accurate (300 energy grid points in total). This is very important in a self-consistent algorithm to ensure converged results for the electron density which has to be integrated over energy (eq. (4.64)). Also shown is the density of states (DOS, blue dashed line). It is obtained by integrating for each lead

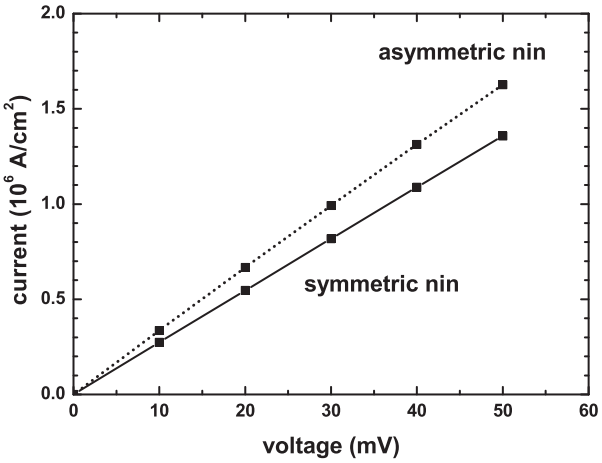


Figure 4.16.: Linear regime of the current–voltage characteristics of the symmetric (solid line) and asymmetric (dotted line) *nin*–structures calculated with the self-consistent CBR method at a temperature of 300 K.

the LDOS over the position, and then adding the contributions of each lead (eq. (4.59)). The DOS has peaks at the onset of the conduction bands edges at the left and right contacts. Note that the energy axis corresponds to the energy E_z along the z direction and not to the total energy E_{total} which includes the integration over $\mathbf{k}_{\|}$. The spin degeneracy factor is included in this figure.

The linear regime of the current–voltage characteristics of the symmetric (solid line) and asymmetric (dotted line) *nin*–structures has been calculated with the self-consistent CBR method and is shown in Fig. 4.16. For the asymmetric n^+in–resistor the applied voltage corresponds to reverse bias operation. In comparison to the symmetric *nin*–structure, the asymmetric resistor shows a higher current density because the effective barrier width and the effective barrier height due to the intrinsic region is reduced. In this small device, the limiting case of ballistic quantum transport is a suitable approximation. For such low biases, the calculated current density of the ballistic calculations do not deviate strongly from calculations that include both scattering and more advanced lead models (not shown). The main reason is the absence of confined states below the conduction band edges of the leads. These states get only filled if scattering is present, and can then influence the charge carrier distribution significantly. In such a case, a ballistic quantum transport model is not adequate. When modeling resonant tunneling devices and especially quantum cascade lasers, it is very important to include scattering [VK10]. The latter might be designed based on e.g. resonant conditions with longitudinal optical phonons (LO phonon scattering).

4.9. Current

The transmission function $T(E)$ can be computed, once the band edge profile of the device has been obtained by means of a charge self-consistent calculation. The ballistic current from lead λ to lead λ' can be calculated based on the Landauer–Büttiker formula (see eq. (4.1)). This equation has to be adjusted for 1D and 2D simulations if the transmission coefficient is a function of E_z or $E_{x,y}$, respectively, rather than of the total energy. In 1D, the integration has to be performed over E_z, and the Fermi function $f(E,\mu)$ has to be replaced with the corresponding Fermi function $f_{1\mathrm{D}}(E_z,\mu)$ given in eq. (4.61). In 2D, the integration has to be performed over $E_{x,y}$ and the appropriate Fermi function $f_{2\mathrm{D}}(E_{x,y},\mu)$ is given in eq. (4.62). The Fermi functions include the corresponding units, so the current in 1D is given in units of $\mathrm{A/m^2}$, in 2D in $\mathrm{A/m}$ and in 3D in A.

If more than two leads are present in the device, then for the total current through a particular lead λ the contributions from all other leads λ' have to be summed up

$$I_\lambda = \sum_{\lambda'=1}^{L} I_{\lambda\lambda'} \qquad (\lambda' \neq \lambda). \tag{4.71}$$

4.10. Conclusions

In this chapter we presented in detail a numerical implementation of the contact block reduction (CBR) method which is a variant of the nonequilibrium Green's function formalism. Charge self-consistent calculations can be performed very efficiently even for 3D structures by means of the CBR approach. Once the potential profile of a device with an arbitrary number of contacts has been obtained, the ballistic current can be calculated based on the Landauer–Büttiker formula. The presented model is suitable for extremely small devices or very low temperatures, where the elastic and inelastic scattering lengths exceed the geometrical device size. In these cases, the ballistic current model is a reasonable model.

5. Calculation of electron mobility in a two-dimensional electron gas

For basic research experiments high-mobility samples are very important where mobilities reach in AlGaAs–GaAs samples up to 10^7 Vs/cm^2. But also for real device applications in e.g. CPUs (transistors), mobile phones (HEMTs) or RF power transistor technology (GaN HEMTs), carrier mobility is a very important parameter that has to be optimized in device design. In particular, the product μn of mobility μ and electron density n is the important parameter because the current is proportional to it. Studying mobility is especially relevant if new materials like InAs or InSb should replace silicon, or if strain is used to enhance the mobility, e.g. uniaxial strain in state-of-the art 22 nm transistors (strained silicon). While the influence of strain on the electron mobility is easier to understand in situations where only the first subband is occupied, the situation becomes much more complicated for holes. This is not only due to the warped and anisotropic hole band structure but also for the fact that more subbands are occupied as the hole energy levels are lying closer to each other. Also for drift–diffusion transport models, the mobility is an important (input) parameter that is essentially a grid point dependent quantity. For silicon channels, mobility models have been developed that require up to 70 parameters (35 for electrons, 35 for holes) [DLP$^+$97]. Such models have their justification in the TCAD industry but are certainly not suited for less well studied materials. For the physicist it is more attractive to rely on models where no mobility parameters are needed as input. It is even better if no (in case of 'ab initio' or 'from first principles' calculations) or only a few input parameters are needed that are experimentally accessible. An example for the latter is the calculation of the transport properties in quantum cascade lasers by the nonequilibrium Green's function (NEGF) method [KYV$^+$09]. However, what we need here within the framework of next**nano** is a simple algorithm that takes the self-consistently calculated band profile of a general two-dimensional electron (2DEG) or hole gas (2DHG), its energy levels (band structure) and wave functions, and their occupation to calculate the mobility. While sophisticated methods have previously been used to calculate the mobility, e.g. by a combination of $\mathbf{k}\cdot\mathbf{p}$ and Monte Carlo methods [OZV98], it seems not to be an easy task to implement such approaches into a general Schrödinger–Poisson solver like next**nano**. In the literature we identified simpler models that might be suitable for us to get rough estimates on the mobility values. One of them is based on a single-band effective mass approach proposed by Shao et al. [SSRM06] that includes scattering from ionized impurities, background neutral impurities, deformation potential acoustic phonons and polar optical phonons to calculate the mobility in delta-doped heterostructures. Later the same authors extended their work to a $\mathbf{k}\cdot\mathbf{p}$ model [SSRMY07] in order to take into account nonparabolicity

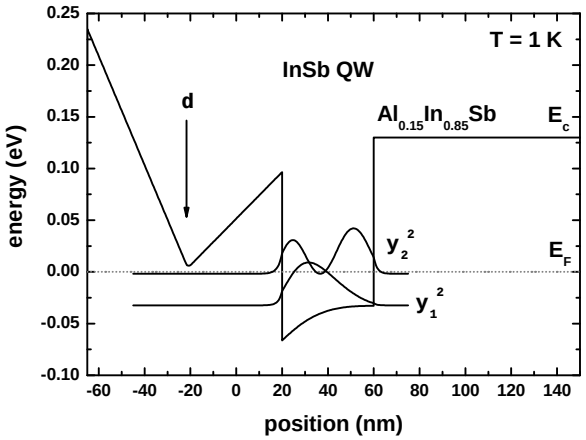


Figure 5.1.: Calculated conduction band edge profile E_c, Fermi level E_F and the square of the lowest two electron wave functions (ψ_1^2, ψ_2^2) at $T = 1\,\mathrm{K}$ of a modulation doped InSb quantum well of width 40 nm.

effects. The latter work is very likely suited to include strain and it might be possible to extend it in order to also calculate hole mobilities. For this reason we implemented their single-band model into next**nano**[3] in order to test it against experimental results. In this thesis, we do not want to reproduce their equations used to calculate the mobility and refer the interested reader to Ref. [SSRM06]. Our implementation is based on these equations with the exception of eq. (3.5) in their paper which we corrected by adding a factor of $\frac{1}{4\pi}$ because of SI units. We only want to mention that we extended their model to include the scattering rate by random alloy scattering (alloy disorder scattering) based on the equations given in Ref. [WRLG84]. In the following we present our results.

Figure 5.1 shows the calculated conduction band edge profile E_c, the Fermi level E_F and the square of the lowest two electron wave functions (ψ_1^2, ψ_2^2) at $T = 1\,\mathrm{K}$ of a modulation doped InSb quantum well (QW) of width 40 nm. It is surrounded by and strained with respect to $Al_{0.15}In_{0.85}Sb$ barriers. At $z = -20\,\mathrm{nm}$ there is a δ-doping layer with a sheet doping density of $1 \cdot 10^{12}\,\mathrm{cm}^{-2}$. It is separated from the InSb QW by a 40 nm $Al_{0.15}In_{0.85}Sb$ spacer layer. This spacer layer is introduced in order to reduce scattering by ionized impurities. We now discuss the agreement and disagreement of the calculated mobility results compared to Fig. 3(a) in Ref. [SSRM06] of Shao et al. for this delta-doped InSb 2DEG channel. Figure 5.2 shows our calculated mobility as a function of temperature. The mobility μ_acoustic due to acoustic phonon scattering is in excellent agreement and also the mobility μ_polar due to polar optical LO phonon scattering (if one takes into account that their figures miss a factor of $\frac{1}{4\pi}$). However, the mobility μ_impurity due to ionized impurity scattering and the mobility $\mu_\mathrm{background}$ due to background impurity scattering differ. It seems that the disagreement is not only due to the different sheet density that has been used. Some minor differences in the input

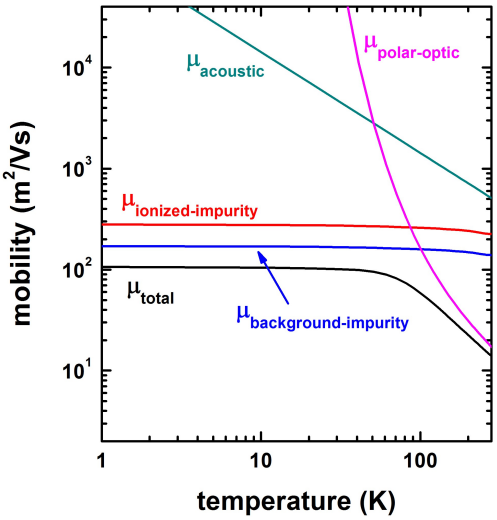


Figure 5.2.: Calculated total electron mobility of the InSb quantum well (black solid line) as a function of temperature. The contributions of the individual scattering mechanisms to the total mobility are indicated by the other solid lines.

parameters are the Schottky barrier height and the conduction band offset. We used $\sim 0.15\,\mathrm{eV}$ whereas Shao et al. used $\sim 0.25\,\mathrm{eV}$. As the implementation of a self-consistent Schrödinger–Poisson algorithm is rather complex, this might be another reason that the results differ slightly.

The mobility is defined as $\mu = e\tau/m$ where m is the effective mass, τ is the scattering time (life time) due to the individual scattering mechanism, and $1/\tau$ is called the transport scattering rate. The total mobility is calculated as usual from the other mobilities by using Matthiessen's rule

$$\frac{1}{\mu_{\mathrm{total}}} = \frac{1}{\mu_{\mathrm{impurity}}} + \frac{1}{\mu_{\mathrm{background}}} + \frac{1}{\mu_{\mathrm{acoustic}}} + \frac{1}{\mu_{\mathrm{polar}}} + \frac{1}{\mu_{\mathrm{alloy}}}. \tag{5.1}$$

To benchmark the implemented model to non-δ-doped structures, other calculations and experiment, we present a further example. Here, we test our algorithm on results of GaAs 2DEGs of another publication. Walukiewicz et al. [WRLG84] calculated the mobility of a modulation doped AlGaAs–GaAs heterostructure and compared their results to the experimental values of Hiyamizu et al. [HSNI83]. We note that our algorithm is suitable for δ-doped QW 2DEGs but this GaAs example is not δ-doped, and not a sharp quantum well but an inversion layer with a triangular confinement potential in the GaAs region. Figure 5.3 shows the calculated conduction band edge profile E_{c}, the Fermi level E_{F} and the square of the lowest two electron wave functions (ψ_1^2, ψ_2^2) at $T = 1\,\mathrm{K}$ of the modulation doped GaAs–AlGaAs heterostructure. The $Al_{0.3}Ga_{0.7}As$ region from 0 nm to 100 nm is doped with a concentration of $1 \cdot 10^{18}\ \mathrm{cm}^{-3}$. A 20 nm $Al_{0.3}Ga_{0.7}As$ spacer layer separates the 2DEG channel from the remote doping area. Figure 5.4 shows our calculated mobility as a function of temperature. The agreement

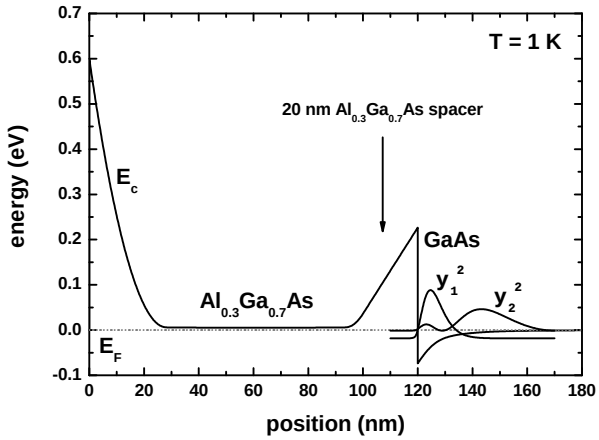


Figure 5.3.: Calculated conduction band edge profile E_c, Fermi level E_F and the square of the lowest two electron wave functions (ψ_1^2, ψ_2^2) at $T = 1\,\mathrm{K}$ of a modulation doped AlGaAs–GaAs heterostructure.

of the total mobility μ_total to experiment (squares) is very good for all temperatures. The experimental values have been extracted from Fig. 5 of Ref. [HSNI83]. At large temperatures the total mobility is dominated by polar optical phonon scattering and for very low temperatures by remote ionized impurity scattering. In the intermediate range, acoustic phonon scattering dominates. Alloy scattering is only relevant for the part of the wave function that penetrates into the AlGaAs barrier. As the material where the 2DEG resides (GaAs) is not an alloy, the mobility due to alloy scattering does not have a dominant influence on the total mobility. Our numerical values for the mobility due to remote ionized impurity scattering at low temperatures is in perfect agreement to the calculations by Walukiewicz (see Fig. 2 in Ref. [WRLG84]) although different models have been used.

We have further tested the algorithm on GaN–AlGaN quantum wells and InGaAs quantum wells where in the latter example the well material is an alloy, and thus the total mobility is dominated by alloy scattering up to temperatures of 100 K (not shown[1]).

The calculated mobility for each temperature step automatically took into account the temperature dependent band gap (eq. (B.2)). The sheet density of the δ-doping layer, the spacer width and the calculated sheet electron density of the 2DEG which also depends on the temperature is used as input for the calculation of the mobility due to remote ionized impurity scattering. The sheet electron density of the 2DEG also enters the mobility due to background impurity scattering, together with the background doping concentration. The quantum well width is input parameter to the mobility due to deformation potential acoustic phonon scattering. In Fig. 5.4 we used $0.35 \cdot 10^{12}\,\mathrm{cm}^{-2}$ as

[1]These results have been documented online. next**nano**^{3} tutorial: Mobility in two-dimensional electron gases (2DEGs), http://www.nextnano.de/nextnano3/tutorial/1Dtutorial_2DEGmobility.htm

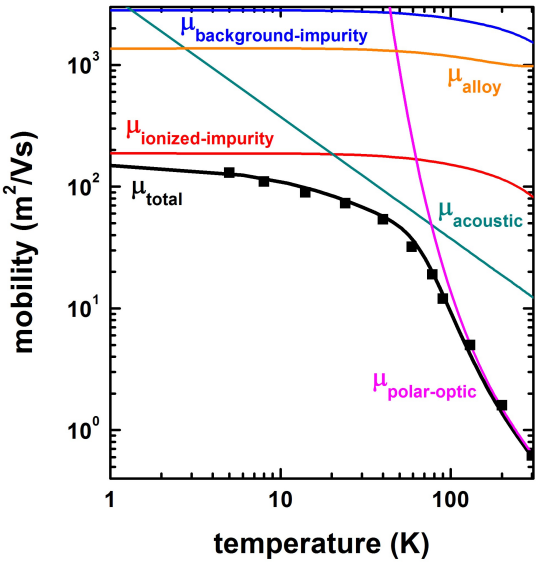


Figure 5.4.: Comparison of calculated total electron mobility μ_{total} of a GaAs 2DEG (black solid line) to experiment (squares) as a function of temperature.

the remote doping sheet density. Furthermore, as there is only a triangular confinement and no defined quantum well width, we approximated the latter input quantity to be 13 nm which corresponds to the extension of the ground state wave function along the growth direction where the effect of barrier penetration has been taken care of. Whereas Walukiewicz did not use the bulk value for the effective electron mass in GaAs but used a higher value of $m_c = 0.076\ m_0$ to take into account nonparabolicity effects, we used $m_c = 0.067\ m_0$ as this gives better agreement to the mobility at higher temperatures and also because this is the usually accepted material parameter for GaAs. In contrast to previous calculations where often for all temperatures the same 2DEG sheet density had been assumed – Walukiewicz used a 2DEG density of $0.3 \cdot 10^{12}$ cm^{-2} – our implemented model allows us to calculate this input parameter dynamically. The background impurity concentration was $9 \cdot 10^{13}$ cm^{-3}. An obvious improvement is the inclusion of interface (surface) roughness scattering, which is important for thin wells and high 2DEG densities, and piezoelectric scattering into eq. (5.1).

In conclusion, we have implemented and evaluated a single-band approach to calculate the electron mobility in two-dimensional electron gases and compared our results to other theoretical calculations and experiment. This allows one to easily identify the relevant scattering mechanisms at various temperature ranges. Device parameters like spacer width, doping concentration, quantum well width or alloy profile can easily be varied to study their influence on mobility, or on the product μn of mobility and electron density in a very transparent way. We conclude that this model seems to be suited to be extended to a $\mathbf{k} \cdot \mathbf{p}$ formalism in order to calculate strained electron or hole mobilities.

Part II.

Biosensors

6. Introduction

In the second part of this thesis I examine semiconductors that are in contact to liquids. Such an arrangement is typical of biosensors where the fields of semiconductor physics, semiconductor technology, electrochemistry, chemistry and biology are involved. Additionally, also expertise in biotechnology or medical research is extremely useful in order to propose and evaluate potential applications. Consequently, a very interdisciplinary collaboration of the experts in each discipline is necessary. In this thesis I contribute my knowledge on modeling of such systems to this very active research area. I focus on the treatment of the semiconductor physics on a quantum mechanical level, and on the description of the electrostatics in the liquid. First, I present a protein sensor based on silicon and describe the modeling approach using the Poisson–Boltzmann equation. I compare this model to the much simpler Debye–Hückel (DH) approximation that was previously used to analyze experimental results. Our calculations show that significant more insight into the actual charge distribution in the electrolyte is obtained which cannot be accounted for within the DH model. Furthermore, I have developed a novel approach to model the charge density profiles at semiconductor–electrolyte interfaces that I will discuss in the following sections. It allows us to distinguish hydrophobic and hydrophilic interfaces. Typically, such a differentiation has not been considered so far in semiconductor based biosensors, mainly because they are made of materials that are passivated by a native oxide layer (e.g. Si–SiO_2 sensors). This insulating region separates the charges in the semiconductor and the charges due to ion accumulation in the electrolyte, thus hiding the importance of the nonpolar or polar character of the interface. I apply this new model to recently developed graphene and diamond based solution gated field-effect transistors. In both materials, no insulating material layer between the semiconductor charge carrier gas and the accumulated ions in the liquid is present. Therefore the influence of the hydrophobic or hydrophilic character leads to completely different results in terms of sensitivity of such devices. By comparing the calculated results to experiment we demonstrate that the hydrophobic character of these materials must be taken into account to properly reproduce and understand experimental results. Our approach extends previous work where potentials of mean force for the ions are included in the Poisson–Boltzmann equation.

7. Theoretical model for the detection of charged proteins with a silicon-on-insulator sensor

The work presented in this chapter is based on the preliminary studies of Uhl [Uhl04]. I have extended his effort and summarized the results in a joint publication [BUBV08] which is presented in this chapter.

For a biosensor device based on a silicon-on-insulator structure, we calculate the sensitivity to specific charge distributions in an electrolyte solution that arise from protein binding to the semiconductor surface. The surface is bio-functionalized with a lipid layer so that proteins can specifically bind to the headgroups of the lipids on the surface. We consider charged proteins such as the green fluorescent protein (GFP) and artificial proteins that consist of a variable number of aspartic acids. Specifically, we calculate self-consistently the spatial charge and electrostatic potential distributions for different ion concentrations in the electrolyte. We fully take into account the quantum mechanical charge density in the semiconductor. We determine the potential change at the binding sites as a function of protein charge and ionic strength. Comparison with experiment is generally very good. Furthermore, we demonstrate the superiority of the full Poisson–Boltzmann equation by comparing its results to the simplified Debye–Hückel approximation.

7.1. Introduction

The quickly progressing technology of low-dimensional semiconductor nanostructures requires and depends on reliable predictive theoretical methods for systematically improving, designing and understanding the electronic and optical properties of such structures. The situation becomes even more complicated if these nanostructures are combined with biomaterials to form biosensors [CWPL01]. These sensors are gaining importance due to their large potential in commercial applications, like pH, protein, virus or DNA sensors (biochips). Ion-selective field-effect transistors (ISFETs) usually contain biomaterials in an electrolyte and consist of a two-dimensional electron (or hole) gas (2DEG) in the semiconductor region where a source–drain voltage is applied in the 2DEG plane perpendicular to the solid–electrolyte interface. The measured source–drain current depends on the electron density in the 2DEG. The goal is to influence the electron density in the 2DEG in a reproducible manner through changes in the electrostatic potential which are caused by the charge distribution inside the electrolyte and in the vicinity of the interface between the semiconductor and the electrolyte, i.e. the electrolyte acts through

this field-effect as a gate. Several variations of this concept are possible, e.g. instead of having a 2DEG one could use a nanowire with quantum confinement in two directions, and thus enhance the sensitivity due to the increased surface-to-volume ratio, or one could use an optical device where the electrostatic potential in the electrolyte modifies transition energies in quantum wells, quantum wires or even quantum dots. Modeling of such devices [HK07] is essential not only for analyzing and interpreting experimental results, but also for verifying theoretical concepts, and for the understanding of how to efficiently improve sensitivity. For a review on the challenges in the use of nanostructures for on-chip biosensing and diagnostics, see e.g. Ref. [Bal10].

In this chapter we present realistic models of the electrolyte solution, its interaction with the semiconductor device surface, and of the semiconductor device itself. We discuss detailed simulations of protein sensors based on silicon in order to demonstrate the applicability of this approach. In Section 7.2, we describe how we solve the Schrödinger equation and calculate the charge density in the semiconductor region. Our method that models the charge density in the electrolyte is outlined in Section 7.3. Comments about the numerical treatment of the coupled system of semiconductor and electrolyte equations are given in Section 7.4. Details of previously performed relevant experiments, and the theoretical model of the sensor are discussed in Section 7.5. In Section 7.6, we present results of self-consistent calculations of the spatial charge and electrostatic potential distributions for various protein charges and different ion concentrations in the electrolyte. Finally, the calculated surface potential is compared with experiments. The results indicate that the full Poisson–Boltzmann equation is able to reproduce experimental data whereas the widely used Debye–Hückel approximation faces severe limitations.

7.2. Modeling the semiconductor

The charge distribution within a general semiconductor device is given by eq. (1.2). The p-type doped silicon sensor that is investigated in this work is operated in the inversion regime. Hence, only a quantum mechanical treatment of the conduction band electrons close to the X points in the Brillouin zone is necessary. The contributions to the density of all other bands like the conduction bands at the Γ and L points, and the heavy, light and split-off hole valence bands are negligible. In bulk silicon there are six equivalent conduction band valleys close to the X points in the Brillouin zone that are described by ellipsoidal effective mass tensors with one longitudinal and two transverse masses, m_l and m_t, respectively. As these ellipsoidal mass tensors are oriented differently with respect to each other, we have to treat these minima separately, i.e. we have to consider three different valleys where each one is twofold degenerate. The quantum mechanical electron charge density for each of these three valleys is given by

$$n(\mathbf{x}) = g_\mathrm{v} g_\mathrm{s} \sum_n |\Psi_n(\mathbf{x})|^2 \, f\left(\frac{E_n - E_\mathrm{F}(\mathbf{x})}{k_\mathrm{B}T}\right), \tag{7.1}$$

where $g_\mathrm{s} = 2$ is the spin degeneracy and $g_\mathrm{v} = 2$ is the valley degeneracy. Ψ_n and E_n are the wave functions and eigenenergies of the three-dimensional Schrödinger equation

(eq. (2.2)) and depend on the orientation of each of the three ellipsoidal mass tensors, i.e. each valley requires the solution of the Schrödinger equation. The occupation of the eigenstates is governed by the Fermi–Dirac distribution function f (eq. (4.63)) taking into account the local quasi-Fermi level $E_{\mathrm{F}}(\mathbf{x})$. k_{B} is Boltzmann's constant and T is the temperature.

We use a standard approach to calculate the energy levels and wave functions, namely the single-band effective mass Schrödinger equation within the envelope function approximation. We discretize this equation with a finite differences method and assume a parabolic energy dispersion [TSCH90]. For a semiconductor structure that is grown along the z direction and that is homogeneous along the x and y directions, the envelope functions $\psi_n(z)$ and the energies E_n of the n quantized electronic states are obtained as the solutions of the one-dimensional Schrödinger equation (eq. (4.14)) where the potential energy is given by $V(z) = E_{\mathrm{c},0}(z) - e\phi(z)$, and $m_{\perp}(z)$ is one of the three effective mass tensor components along the growth direction z, i.e. m_{l} or m_{t} for (001) oriented silicon. $E_{\mathrm{c},0}(z)$ represents the conduction band edge profile of the relevant valley and takes into account band offsets at material interfaces. $\phi(z)$ is the electrostatic potential which is obtained from solving Poisson's equation (eq. (1.1)). It includes the external bias potential and the internal potential resulting from mobile charge carriers and ionized impurities. For a one-dimensional device that is homogeneous along the x and y directions, the quantum mechanical electron charge density is calculated for each valley as

$$n(z) = g_{\mathrm{v}} g_{\mathrm{s}} \sum_n |\psi_n(z)|^2 \, \frac{m_{||}(z) k_{\mathrm{B}} T}{2\pi\hbar^2} \ln\left(1 + \exp\left(\frac{E_{\mathrm{F}}(z) - E_n}{k_{\mathrm{B}} T}\right)\right), \tag{7.2}$$

where the sum over n is only over the lowest occupied subbands. $m_{||}(z)$ is the effective mass in the (x, y) plane. Obviously, this value depends on the conduction band valley, i.e. for unstrained (001) oriented silicon layers, the ground state electron level is associated with the longitudinal electron mass and thus $m_{||}$ is the transverse mass m_{t}. For the other valleys where the transverse mass is oriented along the growth direction z, the parallel mass is calculated as $m_{||} = \sqrt{m_{\mathrm{l}} m_{\mathrm{t}}}$ (density of states mass). Equation (7.2) leads to discontinuous charge densities at material interfaces if the value of $m_{||}$ differs between neighboring materials. In order to avoid this, we calculate for each subband n the parallel mass $m_{||}$ according to Ref. [DT94].

Taking into account the charge neutrality requirement, we first solve the Schrödinger–Poisson equation self-consistently in the whole device with the equilibrium requirement that the Fermi level in the silicon layer is assumed to be constant at $E_{\mathrm{F}} = 0\,\mathrm{eV}$. In this case, we solve the Poisson equation with Neumann boundary conditions and obtain the built-in electrostatic potential. The boundary values of the built-in potential plus optionally applied bias potentials at ohmic or Schottky contacts (Fig. A.1) are then used as Dirichlet boundary conditions for the Poisson equation in nonequilibrium calculations.

7.3. Modeling the electrolyte

An electrolyte is an aqueous solution containing dissolved ions (e.g. Na^+, Cl^-) that result from the dissociation of salts. Electrolytes that are used as biosensors are usually buffer solutions (see Section 8.3) and therefore resist changes in H_3O^+ and OH^- ion concentrations (and consequently the pH) upon addition of small amounts of acid or base, or upon dilution. The concentrations of the ions that are contained in the buffer depend on the pH and the $pK'_{a,T}$ value (dissociation constant) and can be calculated using the well-known Henderson–Hasselbalch equation (eq. (8.12)). In addition, the $pK'_{a,T}$ value depends on temperature and on ionic strength I (eq. (8.9)) in a self-consistent way. For instance, when using a phosphate buffer, the concentrations of the buffer ions at a particular pH are governed by three different $pK'_{a,T}$ values and thus it is extremely difficult to derive the concentrations analytically. However, they can be calculated numerically in an iterative scheme [BE96]. In Section 8.3, we describe the details of our buffer model where we allow the variables pH, $pK'_{a,T}$ and ionic strength to vary with spatial coordinates. Such an approach is necessary for analytes that produce local charge variations in the electrolyte, e.g. a charged molecule that binds to the semiconductor device surface. Furthermore, a local variation of pH is critical for the operation of EnFETs (enzyme field-effect transistors) where the enzyme reaction depends on the pH value.

The distribution of all ion charges in the electrolyte solution is governed by the nonlinear Poisson–Boltzmann equation which is composed of the nonlinear Poisson equation (eq. (1.1)) and the equation that describes the charge density distribution in the electrolyte (eq. (7.3)). Conventionally, the Poisson–Boltzmann equation is linearized which leads to the Debye–Hückel approximation (Section 8.2). However, as we will show in Subsection 7.6.1 such a simplification is generally not applicable in real devices and only valid for special and very limited cases. The solution $\phi(\mathbf{x})$ of the Poisson–Boltzmann equation determines the charge density in the electrolyte at position $\mathbf{x}$

$$\rho(\mathbf{x}) = \sum_{i=1}^{N} z_i e c_i(\mathbf{x}) = \sum_{i=1}^{N} z_i e c_{i,0} \exp\left(-\frac{z_i e\left(\phi\left(\mathbf{x}\right) - U_\mathrm{G}\right)}{k_\mathrm{B} T}\right), \tag{7.3}$$

where z_i is the ion valency, e is the positive elementary charge, c_i is the resulting ion concentration and $c_{i,0}$ is the bulk concentration of the ion species i. The bulk electrolyte potential $\phi(\infty)$ can be adjusted by varying the potential of the reference gate electrode U_G that is connected to the electrolyte (Dirichlet boundary condition). $\phi(\mathbf{x})$ is the electrostatic potential that is obtained by solving the nonlinear Poisson equation in the overall device self-consistently and $k_\mathrm{B}T$ is the thermal energy of the system. Interface reactions can be taken into account by the so-called site-binding model for amphoteric oxide surfaces [Ber70, HW78] where the adsorption and dissociation of H^+ and OH^- ions at the interface between the electrolyte and the oxide lead to interface charge densities which depend on both the electrostatic potential at the interface and the pH of the electrolyte. These interface charge densities simply have to be added to the charge density that enters the Poisson equation.

7.4. Modeling the coupled system of semiconductor and electrolyte

The electrostatics within the electrolyte and the semiconductor require the self-consistent solution of the Poisson and Schrödinger equations. Both equations are discretized on a nonuniform grid with a finite differences method. They are solved numerically by iterative methods that are described in more detail in Ref. [TZA+06]. We point out that we solve only one single Poisson equation, given in eq. (1.1), which includes both the electrolyte as well as the semiconductor region. ε_r is then the static dielectric constant of either the electrolyte or any of the semiconductor or insulator materials. In regions where the electrolyte is present, the charge density $\rho(\mathbf{x})$ is described by eq. (7.3), and in regions where the semiconductor materials or the oxides are present, the typical semiconductor equation is used (eq. (1.2)) which may include a suitable fixed or variable sheet charge density at the interface between the semiconductor device and the electrolyte. We note that it is not necessary to solve the Schrödinger equation in regions where the quantum mechanical density is negligible or zero, e.g. in insulators. However, wave function penetration into the barrier materials (e.g. at Si–SiO_2 interfaces) is fully taken into account by including a small region of the barrier material into the Schrödinger equation. We have implemented the above mentioned equations and similar ones for the two- and three-dimensional Schrödinger equations into the software package next**nano**^{3} [www]. This enables us to model combined semiconductor–electrolyte systems in one [BUV05], two [Ped06] and three dimensions for arbitrary geometries and material compositions. Typically, nanowire sensors are more sensitive than planar sensors, and if the nanowire dimensions are less than 50 nm, the sensitivity can increase even further [Ped06].

7.5. Description of the geometry and composition of the protein sensor

7.5.1. Sensor structure

Here, we discuss a silicon-on-insulator (SOI) based thin-film resistor that we will model in detail in Section 7.6. Indeed, such a device has been realized experimentally for chemical and biological sensor applications [NRL+03, NRB04]. Peptides with a single charge can be detected and it is possible to distinguish single charge variations of the analytes even in physiological electrolyte solutions [LNH+06].

Figure 7.1 shows the layout of this bio-functionalized silicon-on-insulator device. It consists of a SiO_2–Si–SiO_2 structure. Specifically, we take a silicon dioxide buffer layer with a thickness of 200 nm and a conducting silicon layer of 30 nm which is homogeneously p-type doped with boron (doping density $p = 1 \cdot 10^{16}$ cm^{-3}). The silicon layer is covered by a native SiO_2 layer with a thickness of 2 nm. This oxide layer is passivated by an ODTMS (octadecyltrimethoxysilane) monolayer which is required for the bio-functionalization of the semiconductor device. We take a 1.5 nm thick oxide-like ODTMS layer and use a static dielectric constant of $\varepsilon_r = 1.5$. Due to the passivation

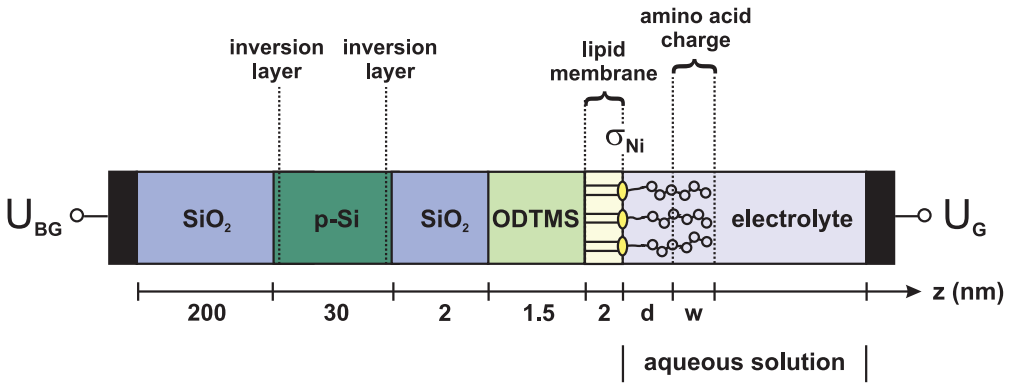


Figure 7.1.: Schematic layout of the considered SOI structure (after C. Uhl [Uhl04]). There is a negative interface charge density σ_{Ni} at the lipid–electrolyte interface. The amino acid charge is assumed to be distributed homogeneously over a width w. The electrolyte region includes the histidine-tagged amino acids as well as the neutral part of the tag of length d.

by ODTMS, we assume that no interface charges are present at the native oxide surface. The ODTMS layer is surface-functionalized with a lipid membrane that allows for the specific binding of molecules. This lipid monolayer (2 nm) consists of DOGS-NTA (1,2-dioleoyl-*sn*-glycero-3-{[*N*(5-amino-1-carboxypentyl)iminodiacetic acid]succinyl}) incorporated into two matrix lipids (DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine) and cholesterol). The lipid membrane is treated as an insulator using the same material parameters as for ODTMS. Thus, no charge carriers are assumed to be present within this layer. As the lipid layer is very dense, no electrolyte is considered within the lipid region.

For the ionic content of the electrolyte we consider a variable concentration of KCl (10, 50, 90 or 140 mM), and a fixed concentration of 1 mM of $NiCl_2$ and 1 mM of phosphate buffer saline (PBS, see Subsection 8.3.5) solution, respectively. The $NiCl_2$ dissociates into 1 mM of doubly charged cations and 2 mM of singly charged anions. For all calculations, the pH of the bulk electrolyte has been set to 7.5. The calculated concentrations of the PBS buffer ions are listed in Table 7.1 for different salt concentrations. These values refer to the bulk electrolyte. In the vicinity of the semiconductor surface and in regions around charged analytes, however, the actual concentrations of the buffer ions vary locally. Our buffer model automatically takes this into account because the spatial variations of pH, ionic strength and $\mathrm{p}K'_{\mathrm{a},T}$ are determined self-consistently (see Section 8.3 for more details). The ionic strengths of the electrolyte solutions considered in this work are largely dominated by the respective concentrations of singly charged anions and cations from KCl as can be seen in Table 7.1. In these particular cases, i.e. small concentrations of PBS with respect to KCl, the Debye screening length κ^{-1} (eq. (8.5)) is fully dominated by the KCl concentration.

The functionalized surface exposes NTA headgroups that carry two negative charges to the electrolyte solution. They have the ability to form a chelate complex with nickel ions

Table 7.1.: Concentrations of ions in units of mM, ionic strength I and Debye screening length κ^{-1} for several configurations of the electrolyte (1 mM PBS, 0 or 1 mM $NiCl_2$, pH = 7.5, $T = 25\ ^\circ C$)

Ion	0 mM KCl	10 mM KCl	50 mM KCl	90 mM KCl	140 mM KCl
$[H_2PO_4^-]$	0.303	0.256	0.214	0.192	0.176
$[HPO_4^{2-}]$	0.697	0.740	0.786	0.808	0.824
$[PO_4^{3-}]$	$0.135 \cdot 10^{-4}$	$0.206 \cdot 10^{-4}$	$0.335 \cdot 10^{-4}$	$0.430 \cdot 10^{-4}$	$0.524 \cdot 10^{-4}$
$[Na^+]$	1.697	1.740	1.786	1.808	1.824
$[K^+]$	0	10	50	90	140
$[Cl^-]$	0	10	50	90	140
$[Ni^{2+}]$	0	1	1	1	1
$[Cl^-]$	0	2	2	2	2
$[H^+]$	$0.316 \cdot 10^{-4}$	$0.316 \cdot 10^{-4}$	$0.316 \cdot 10^{-4}$	$0.316 \cdot 10^{-4}$	$0.316 \cdot 10^{-4}$
$[OH^-]$	$0.316 \cdot 10^{-3}$	$0.316 \cdot 10^{-3}$	$0.316 \cdot 10^{-3}$	$0.316 \cdot 10^{-3}$	$0.316 \cdot 10^{-3}$
I (mM)	2.393	15.481	55.573	95.616	145.648
$\kappa\ (\mathrm{nm}^{-1})$	0.159	0.405	0.768	1.007	1.243
κ^{-1} (nm)	6.277	2.468	1.302	0.993	0.805

if the latter are present in the solution. Upon loading with nickel, the charge of the headgroup changes by $+1e$ [SDT94] and is then considered to be $-1e$. This results in a negative sheet charge density σ_{Ni} at the lipid–electrolyte interface. The surface density of the DOGS-NTA lipids is considered to be 5% ($f_{\mathrm{NTA}} = 0.05$). The approximated headgroup area, i.e. the average area per functional DOGS-NTA is assumed to be $A_{\mathrm{NTA}} = 0.65\,\mathrm{nm}^2$. Consequently, the density of the headgroups is $s_{\mathrm{NTA}} = f_{\mathrm{NTA}}/A_{\mathrm{NTA}} = 7.7 \cdot 10^{12}\ \mathrm{cm}^{-2}$, so that the resulting charge density σ_{Ni} is given by $\sigma_{\mathrm{Ni}} = -e s_{\mathrm{NTA}}$, where we have assumed that each headgroup carries one negative charge upon exposure to Ni.

The charge carrier concentration of the conducting silicon layer is controlled by applying a back gate voltage U_{BG} which allows for switching between the accumulation (negative U_{BG}) and the inversion regime (positive U_{BG}) and particularly for tuning the sensitivity of the device. In all calculations that are mentioned in this work, the sensor is operated in inversion at $U_{\mathrm{BG}} = 25\,\mathrm{V}$ to allow for comparison with experiment. The experiment has shown that this is a compromise between the highest possible sensor signal and a low noise level. With this configuration, the p-doped silicon channel is inverted and becomes n-type. One can adjust the potential in the electrolyte by varying the voltage U_{G} of the reference electrode in the solution, mainly to vary the electron density of the right inversion layer with respect to the left one (see Fig. 7.1 and Fig. 7.2). In order to get a reasonable magnitude for the charge density in the right channel close to the functionalized surface, we have set $U_{\mathrm{G}} = 1.0\,\mathrm{V}$ for all calculations. These assumptions allow for a realistic comparison with specific protein binding experiments [LNH+06].

7.5.2. Model of the protein charge distribution

We consider two types of proteins: Aspartic acids and the green fluorescent protein. If divalent nickel ions (Ni^{2+}) are bound to the NTA headgroups of the lipid membrane, this surface functionalization then allows for the specific coupling of histidine-tagged (his-tag) proteins or peptides to the membrane [SDT94]. This process can be reversed by adding EDTA (ethylenediaminetetraacetic acid) to the electrolyte. A his-tag is a short amino acid sequence including histidines. They can be fused to one end of a protein and can also bind transition metal cations. We study a protein charge distribution that is spatially separated from the lipid membrane due to a neutral tag of width d in between the charged protein and the lipids. For simplicity, the protein charge is assumed to be distributed homogeneously over a width w.

Aspartic acid

We consider an artificial protein structure where amino acids are tagged to a histidine chain. This artificial peptide binds to an NTA headgroup of the lipid membrane. A part of this artificial protein remains uncharged since no amino acids get attached there. By contrast, the rest of the histidine backbone is negatively charged since we consider aspartic acids that carry one negative charge each for the binding to the tag. It is possible to manufacture the hexahistidine-tagged (His6) peptides with different numbers of charged residues, i.e. one can engineer the number of aspartic acids (Asp) that bind to the tag. The charge of the aspartic acids have been varied between carrying a single charge (His6Asp1) and up to ten charges (His6Asp10). It is expected that for each charge, a different signal can be detected and that peptides with higher charges result in an increased sensor response. The width of the neutral part has been taken to be $d = 2.3\,\mathrm{nm}$ or $d = 2.8\,\mathrm{nm}$, depending on the concentration of KCl. The length of the uncharged part of the peptide consists of the length of the complete NTA headgroup including the spacer of 12 carbon atoms plus the his-tag. The width of the charged part

$$w(n) = n \cdot b \tag{7.4}$$

has been assumed to depend linearly on the number n of aspartic acid units and on the length b of one aspartic acid residue [$LNH^{+}06$]. Thus, the spatial extent of the charge density increases with the number of aspartic acids. Each additional aspartic acid therefore shifts the center of the charge distribution about $b/2$ farther away from the lipid membrane. Additionally, we perform calculations where we keep this width w constant. The integrated charge density in the protein region changes in magnitudes of $-es_{\mathrm{NTA}}$ by increasing the number n of the aspartic acid units. Furthermore, we assume that the amino acid charges tend to repel each other. It is plausible that the strength of this repulsion is influenced by the ionic strength of the electrolyte. For that reason we reduce the length b of one aspartic acid unit at large ion concentrations. The electrolyte region starts at the membrane surface and includes the regions of both the neutral part of the tag and the protein charge distribution so that the ions in the aqueous solution screen the protein charge (Fig. 7.1).

Green fluorescent protein

As a second protein, we consider the binding of the so-called green fluorescent protein[1] (GFP) to the lipid membrane. GFP is also histidine-tagged to the NTA headgroups of the membrane. The size of GFP is larger (length of 4-5 nm) compared to his-tagged aspartic acids. We assume a charge distribution of width $w = 3.0\,\mathrm{nm}$ that is connected with a neutral tag of width $d = 2.3\,\mathrm{nm}$ to the NTA headgroups. At pH = 7.5, GFP carries eight negative charges that we assume to be homogeneously distributed over the protein region w.

7.6. Results of the calculations

Since we have specified all about the sensor and the proteins in the electrolyte, we are now ready to calculate the electrostatic potential in the semiconductor–electrolyte system for several protein charge distributions. The quantum mechanical charge densities are calculated self-consistently by solving the Schrödinger equation in the silicon channel. The Schrödinger and Poisson equations are coupled via the electrostatic potential and the charge densities.

First, we estimate the change in surface potential ϕ_s when one loads the NTA lipids with Ni^{2+}. We assume a sheet charge density change of $\Delta\sigma = -2es_\mathrm{NTA} - \sigma_\mathrm{Ni}$. We have actually calculated that the surface potential increases for a 140 mM KCl solution by 13.5 mV which is in agreement with the measurements [LNH$^+$06].

Figure 7.2 shows the calculated conduction band edge and the electron density in the silicon channel for a back gate voltage of $U_\mathrm{BG} = 25\,\mathrm{V}$. Indicated is also the position of the Fermi level E_F and the electrostatic potential. Specifying a value for the potential U_G of the reference electrode is equivalent to a Dirichlet boundary condition for the electrostatic potential of the Poisson–Boltzmann equation. An increase of U_G leads to higher electron densities in the right channel. Therefore, the variation of U_G and the back gate voltage U_BG allows one to increase the sensitivity of the sensor by adjusting the ratio of the densities of the two channels. Our calculations yield channel densities of the order of a few $10^{12}\ \mathrm{cm}^{-2}$. They are modulated slightly by the actual configuration of the system in terms of ion concentrations and protein charges. Since a lower surface potential ϕ_s yields a lower electron density in the inverted silicon channel, the source–drain current is expected to decrease if negatively charged proteins bind to the functionalized sensor surface.

7.6.1. Influence of the protein charge on the sensitivity

In this subsection we discuss results of the artificial protein that consists of several aspartic acids as described in Subsection 7.5.2. For the 50 mM KCl solution, the neutral part of the histidine tag is assumed to have a width $d = 2.8\,\mathrm{nm}$ and the respective protein charges are homogeneously distributed over a distance $w = nb$ where n is the number

[1] In 2008, the Nobel prize in chemistry was jointly awarded to Shimomura, Chalfie and Tsien "for the discovery and development of the green fluorescent protein, GFP".

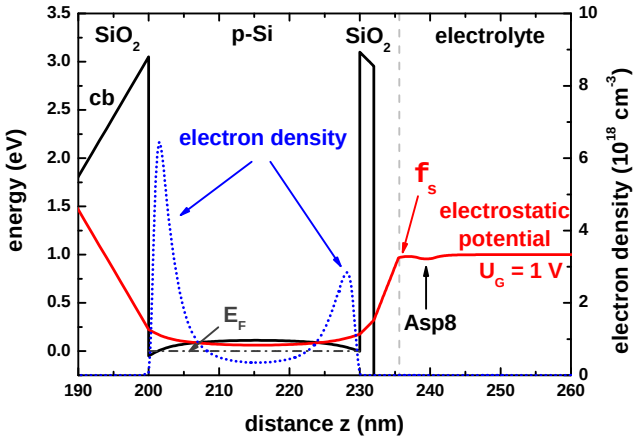


Figure 7.2.: Calculated conduction band edge (black solid line) and electrostatic potential (red solid line) for the SOI structure at 50 mM KCl. The electron charge density (blue dotted line) of the two inversion layers is shown. In the semiconductor, the Fermi level (gray dash-dotted line) is set constant at $E_F = 0\,\text{eV}$. Upon binding to the lipid membrane, the charge of the aspartic acid (Asp8) modifies the surface potential ϕ_s. The interface between the lipid membrane and the electrolyte is indicated by the vertical dashed line.

of aspartic acid units and $b = 0.3\,\text{nm}$. For the 140 mM KCl solution, the respective values are $d = 2.3\,\text{nm}$ and $b = 0.1\,\text{nm}$ [LNH$^+$06]. These parameters are reasonably close to the chemical structure of the histidine-tagged amino acids. Figure 7.3 shows the calculated potential distributions for a varying number of aspartic acids at 50 mM KCl. The magnitude of the negative protein charge density increases with the number of aspartic acids. This results in a lower electrostatic potential in the protein region. Also, the surface potential ϕ_s decreases with increasing protein charge. The region of the charged part of this protein is indicated schematically by the shaded triangle. We note that the electrolyte region starts at the lipid surface at 235.5 nm.

In the following, we calculate the potential change at the interface between the lipid membrane and the electrolyte as a function of the number of aspartic acids that are attached to each histidine tag for KCl concentrations of 50 mM and 140 mM. The reference level ϕ_s^{ref} for the scale of the surface potential change is set to the case for zero protein charge. The surface potential change $\Delta\phi_s$ is then defined as

$$\Delta\phi_s(n) = \phi_s^{\text{ref}} - \phi_s(n), \tag{7.5}$$

where n denotes the number of aspartic acid units. A positive potential change therefore implies that the reference level is higher compared to the situation with a nonzero number of aspartic acids. The results are shown in Fig. 7.4 and show excellent agreement with the experimental data of Ref. [LNH$^+$06] where the surface potential was extracted

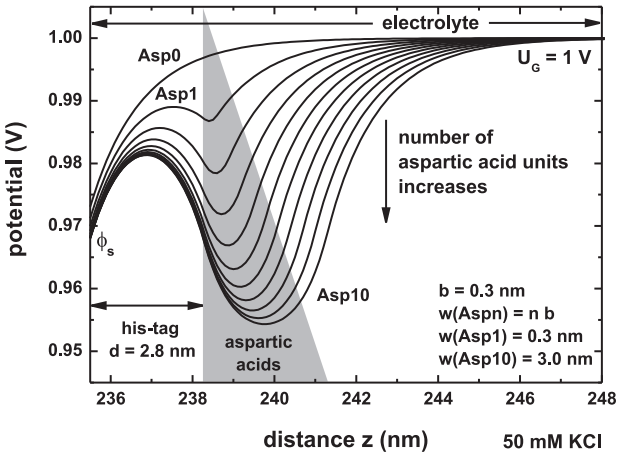


Figure 7.3.: Calculated electrostatic potential distributions for varying protein charge at 50 mM KCl. Shown are the cases where no acids are bound and where the number of acids is $n = 1, 2, 3, ..., 10$. From top to bottom, the number of aspartic acids increases. The width w of the negative protein charge distribution is assumed to increase linearly with the number of aspartic acids. This width is indicated schematically by the shaded triangle.

from measurements of the sheet resistance of the silicon channel. Due to the lower ion concentrations in the case of 50 mM KCl, the protein charge density is less efficiently screened. Consequently, the surface potential change is larger compared to the case of 140 mM KCl. Therefore, the variation of the charge density in the silicon channel – and thus the sensitivity – is greater for a 50 mM KCl solution, as compared to 140 mM KCl.

One important parameter of our model is the width d of the neutral part of the histidine tag. This distance between the lipid membrane and the beginning of the charge distribution of the aspartic acids influences the screening of the protein charges by ions in the solution. Hence, the impact of the amino acid charges decreases with increasing spacing of the lipid membrane, assuming the same protein charge distribution. This means that the influence on the semiconductor device can be enhanced by using a tag that allows small distances of the protein to the lipid membrane.

So far, we have considered homogeneous protein charge distributions where the width w has been varied as a function of the number of aspartic acid units n. In the following, we demonstrate that also a constant width w reproduces experimental data. Now, the number of aspartic acids solely determines the magnitude of the charge density but leaves the spatial extent of the peptide unchanged. For the 50 mM KCl solution, we use $d = 2.8$ nm and a constant width of $w = 1.5$ nm for all n. For the 140 mM KCl concentration, values of $d = 2.3$ nm and $w = 0.5$ nm are taken. The surface potential change as a function of the number of aspartic acids is illustrated in Fig. 7.5 for a protein charge of constant width w. Indicated are the results for both the Poisson–Boltzmann (PB) and the Debye–Hückel (DH) equation (Section 8.2). The latter show a linear

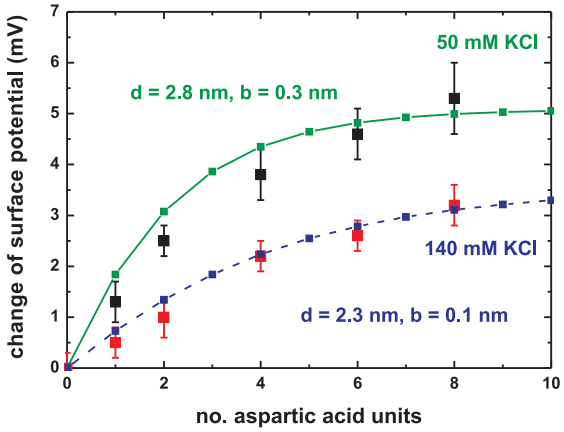


Figure 7.4.: Calculated surface potential change at the lipid membrane as a function of the number of aspartic acids in the artificial protein. The solid line depicts the case of a KCl concentration of 50 mM whereas the dashed line represents the case of 140 mM KCl in the electrolyte solution. The experimental data points are from Ref. [LNH$^+$06].

variation of the surface potential change with the number of aspartic acids. This is expected from the DH equation because a linear variation of the charge density leads to a linear variation of the surface potential change. In contrast, the results of the nonlinear PB equation resemble a logarithmic behavior and are in very good agreement with the experimental data. This nonlinear dependence is attributed to screening effects in the electrolyte which cannot be reproduced correctly within the DH approximation. It is important to note that both assumptions, i.e. a constant width w and the linear variation of w with the number of aspartic acids n, reproduce experimental data, whereas the DH equation leads to unsatisfactory results. This emphasizes the importance of using the full PB equation rather than the linearized DH equation as it allows more insight into the screening of charges in electrolyte solutions. This is especially true if complex bio-functionalized surfaces are used where the binding of charged molecules occurs at about 5-10 nm from the surface, and where the prediction of the sensitivity limitations is desirable.

The differences in the surface potential change that is either obtained within the full Poisson–Boltzmann theory or the simplified Debye–Hückel approximation can be further understood by investigating the spatial potential distributions in the electrolyte for different numbers of aspartic acids. This is shown in Fig. 7.6 where part (a) refers to the PB solutions and part (b) depicts the solutions of the DH approximation. Again, we have included the cases for integer numbers of aspartic acids n from 0 to 10. In both figures, the number of aspartic acids increases from the top to the bottom. As one can see from Fig. 7.6(b), the potential differences at the membrane surface between adjacent potential solutions are constant. By contrast, the potential solutions in Fig. 7.6(a) do not show this behavior, in agreement with the experiment (see also Fig. 7.5).

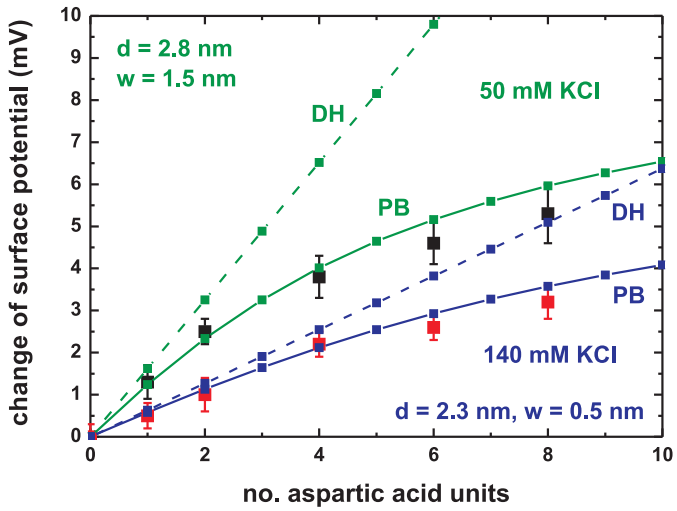


Figure 7.5.: Calculated surface potential change as a function of the number of aspartic acids for two different salt concentrations (50 mM and 140 mM). Included are the results for the solution of the Poisson–Boltzmann (PB) equation (solid lines) and the Debye–Hückel (DH) equation (dashed lines). The latter show a linear dependence and deviate substantially from the PB results which are in good agreement with the experimental data.

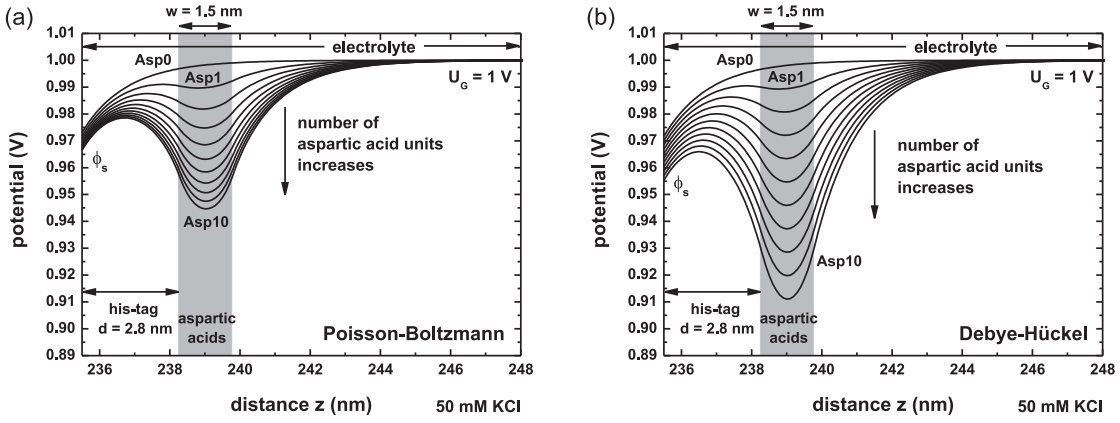


Figure 7.6.: Calculated spatial potential distributions for (a) the full Poisson–Boltzmann equation and (b) the linearized Debye–Hückel equation for a negative protein charge distribution of constant width $w = 1.5\,\mathrm{nm}$. This width is indicated by the shaded region. Both figures include the solutions for different numbers of aspartic acids $n = 0, 1, ..., 10$.

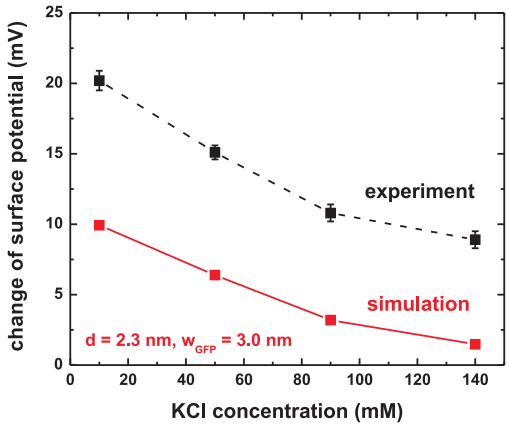


Figure 7.7.: Calculated change of the surface potential as a function of the KCl concentration where we have assumed a charge distribution that resembles the green fluorescent protein. Included are experimental data of Ref. [LNH$^+$06]. The lines are a guide to the eye.

7.6.2. Influence of the ionic strength on the sensitivity

In this subsection we use the same sensor structure as in Fig. 7.1 but detect another protein. We consider the specific binding of the green fluorescent protein (GFP) to the lipid membrane and calculate the change of the surface potential as a function of the salt concentration (10, 50, 90 or 140 mM KCl) in the electrolyte. At pH = 7.5, GFP carries a net negative charge of $-8e$ as can be derived from the primary structure if one calculates the charge of the side chains for the used buffer solution. Consequently, the integrated charge density of GFP is given by $\sigma_{\mathrm{GFP}} = -8e s_{\mathrm{NTA}}$. We assume for simplicity that this charge is distributed evenly over a distance of $w_{\mathrm{GFP}} = 3\,\mathrm{nm}$ which is close to the length of the GFP (4-5 nm). Based on the previous section, the length of the neutral part of the tag has been taken to be $d = 2.3\,\mathrm{nm}$. Here, this length has been assumed to be the same for all KCl concentrations.

We have calculated the change of the surface potential as a function of the KCl concentration in the electrolyte. The electrolyte has the same properties as for the aspartic acids (1 mM PBS, 1 mM $NiCl_2$). Figure 7.7 shows the results and compares them to the experimental data of Ref. [LNH$^+$06]. The trend of the influence of the ionic strength on the sensitivity is well reproduced by our calculations. We note that the exact orientation of the GFP molecule at the surface is not known. A slight tilt angle can increase the measured sensor response due to the exponential dependence of the signal from the distance of the charges. The reduction of the spacing d or the width w_{GFP} of the charge distribution leads to larger surface potential changes. At higher ion concentrations, the Debye screening length of the electrolyte decreases, and thus the charges of the protein are more efficiently screened by the ions in the electrolyte. This leads to a reduced sensitivity which is approximately linear to the inverse of the Debye screening length. The Debye screening lengths of the different salt concentrations are listed in Table 7.1.

Table 7.2.: Material parameters

Description	Symbol	Value	Units
longitudinal electron effective mass (Si)	m_l	0.916	m_0
transverse electron effective mass (Si)	m_t	0.190	m_0
static dielectric constant (Si)	ε_r	11.7	
static dielectric constant (SiO_2)	ε_r	3.8	
static dielectric constant (ODTMS)	ε_r	1.5	
static dielectric constant (electrolyte)	ε_r	80	

7.7. Conclusions

In this section we presented calculations on the sensitivity of a silicon-on-insulator structure with respect to specific charge distributions in the electrolyte solution that may arise from protein binding to the semiconductor surface. Screening effects in the electrolyte have been taken into account using the Poisson–Boltzmann equation. The potential change at the bio-functionalized semiconductor surface has been calculated for various protein charge distributions. Comparison with experiment is generally very good. We have demonstrated the superiority of the Poisson–Boltzmann equation by comparing its results to the simplified Debye–Hückel approximation. In agreement with experiment, we have found that the sensitivity of the structure is enhanced at low ion concentrations. We demonstrated that our numerical approach – the self-consistent solution of the Schrödinger and Poisson–Boltzmann equation – is well suited to model semiconductor based biosensors in a systematic manner, which is a requirement in order to both understand and optimize their sensitivity. Relevant material parameters that were used in the calculations of this section are listed in Table 7.2.

8. Extension to the Poisson–Boltzmann equation

In this chapter we first discuss an analytical solution to the Poisson–Boltzmann equation – the Gouy–Chapman solution. Then we briefly show how the Poisson–Boltzmann equation can be linearized. This is known as the Debye–Hückel approximation. We then summarize the equations that we used in the previous chapter to include buffer solutions in our algorithm. These three sections are closely related to the previous chapter on silicon based protein sensors. Finally we discuss how the Poisson–Boltzmann equation can be extended to include potentials of mean force (PMF) for the ions. Essentially, these PMFs modify the concentrations of different ions in the vicinity of a surface. These potentials are different for each ion type, and they depend on the properties of the surface, e.g. if it is hydrophobic or hydrophilic. The latter also influences the water density close to the surface which will be taken care of within this model. We call this model the extended Poisson–Boltzmann equation and it leads at the surface to a significantly different ion distribution in the electrolyte compared to the standard Poisson–Boltzmann model. This will be demonstrated in both cases for a simple monovalent salt. For the Gouy–Chapman solution the results of a NaCl or NaI salt will be identical as only the charge and the valency of the ions are input to the equation. In the extended Poisson–Boltzmann model the results for NaCl and NaI will differ as the PMFs are different for Cl^- and I^-. The extended Poisson–Boltzmann model will be the topic of the following two chapters on graphene and diamond based solution gated field-effect transistors.

8.1. The Gouy–Chapman solution

In general, there are only very rare cases where the Poisson–Boltzmann (PB) equation can be solved analytically. The Gouy–Chapman solution is one example of such a solution. It is valid for a planar solid–electrolyte interface and for a symmetric salt like NaCl which consists of singly charged cations (Na^+) and singly charged anions (Cl^-) in a solution. The Gouy–Chapman solution relates the surface charge σ_s at the interface to a position dependent electrostatic potential $\phi(x)$ in the electrolyte which determines the ion distribution. The surface potential ϕ_s is related to the surface charge density via the Grahame equation which is also valid for nonsymmetric salts like divalent $CaCl_2$. For a monovalent salt it is particularly simple and reads

$$\sigma_\mathrm{s} = \sqrt{8\varepsilon_0\varepsilon_\mathrm{r}k_\mathrm{B}T}\sinh\left(\frac{e\phi_\mathrm{s}}{2k_\mathrm{B}T}\right)\sqrt{c_\mathrm{salt}}, \tag{8.1}$$

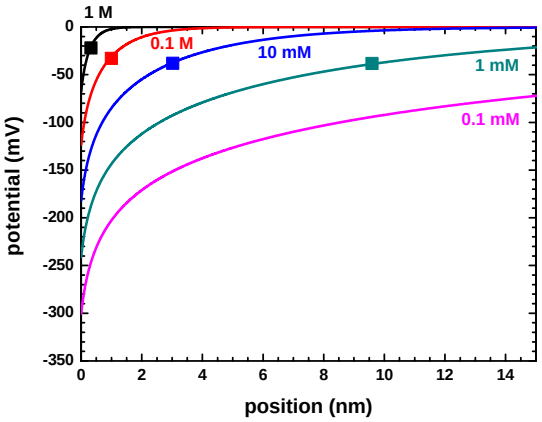


Figure 8.1.: Calculated electrostatic potential distribution (Gouy–Chapman solution) for different salt concentrations of NaCl at a fixed surface charge at the solid–electrolyte interface of $\sigma_s = -0.2$ C/m^2. The squares indicate the Debye screening lengths.

where c_{salt} is the salt concentration. In the following, we show our numerical solution of the PB equation for a monovalent salt which is equivalent to the Gouy–Chapman solution. The NaCl concentration is varied from 0.1 mM to 1 M[1] at a temperature of $T = 25$ °C. The static dielectric constant of water is assumed to be $\varepsilon_r = 78$. Figure 8.1 shows the electrostatic potential for different salt concentrations at a fixed surface charge at the solid–electrolyte interface of $\sigma_s = -0.2$ C/m^2 $= -124.8 \cdot 10^{12}$ |e|/cm^2 as a function of distance from the interface. The squares indicate for each salt concentration the calculated values of the Debye screening length: 0.3 nm for the 1 M, 3 nm for the 10 mM and 31 nm for the 0.1 mM NaCl solution. The Debye screening length is defined in eq. (8.5) and is shown as a function of concentration for a monovalent salt such as NaCl in Fig. 8.2. For a monovalent salt the nominal value of the salt concentration is equal to the ionic strength (eq. (8.9)) which is a measure for the screening of charges in a solution.

Figure 8.3 shows the resulting ion distribution for the 0.1 M NaCl electrolyte. The multiples of the Debye screening lengths are indicated by the vertical lines. The negative surface charge is screened by the positive Na^+ ions (solid line) which are attracted to the surface whereas the negatively charged Cl^- ions (dotted line) are repelled from the surface. At about 5 nm, both ions again reach their equilibrium distribution of 0.1 M. One can see a clear weakness of the Poisson–Boltzmann equation, namely that ions are allowed to come infinitely close to the interface where they reach very high concentrations.

The Gouy–Chapman solution can also be used to assess the linearization of the

[1]The SI-unit for the molar concentration (or molarity) is mol/m^3. However, typically the unit 'molar' (M = mole/liter) is used. 1 M = 1 mol/l = 1000 mol/m^3

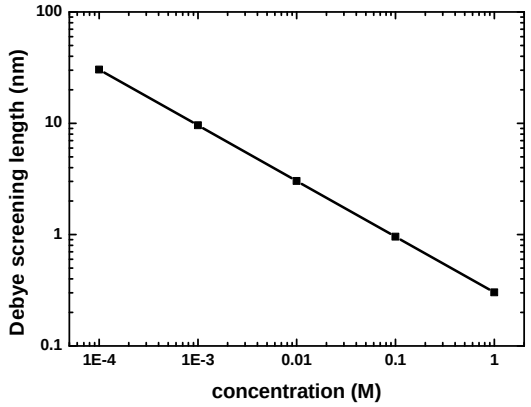


Figure 8.2.: Debye screening length as a function of salt concentration for a monovalent salt such as NaCl showing typical length scales

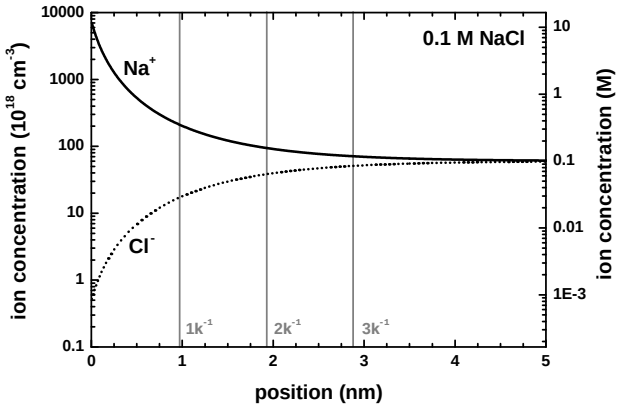


Figure 8.3.: Distribution of Na^+ and Cl^- ions as a function of distance from a negatively charged solid–electrolyte interface. The Debye length κ^{-1}, and multiples of it, are indicated by the vertical lines. The equilibrium density of 0.1 M is reached after 5 nm.

Poisson–Boltzmann equation (Debye–Hückel approximation, Section 8.2). In this approximation, surface charge σ_s and surface potential ϕ_s can be related through the basic capacitor equation

$$\sigma_s = \phi_s C_{DL}, \tag{8.2}$$

where C_{DL} is the capacitance per unit area of the electric double layer. This approximation is only valid for low potentials (up to several tens of mV) where the surface charge density is proportional to the surface potential. The validity of this approximation is

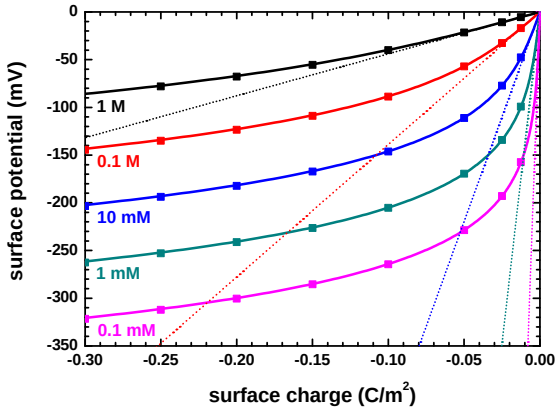


Figure 8.4.: Calculated surface potential (squares) as a function of surface charge for NaCl at different salt concentrations. The solid lines show the analytical solution of the Grahame equation for a monovalent salt whereas the dotted lines are the solutions of the Debye–Hückel approximation. Only for high salt concentrations or small surface charges, the linear relation of surface charge and surface potential is fulfilled.

better for higher salt concentrations.

Figure 8.4 shows the surface potential at the solid–liquid interface as a function of interface charge for the monovalent salt NaCl at different salt concentrations calculated numerically with the Poisson–Boltzmann equation (squares). The solid lines are the analytical solutions of the Grahame equation (eq. (8.1)) for a monovalent salt demonstrating that our calculations are correct. The dotted lines have been calculated using the Debye–Hückel approximation (eq. (8.7)). It can be clearly seen that only for high salt concentrations or small surface charges the linearization is valid. In Fig. 8.4, for 0.1 M NaCl, $\sigma_\mathrm{s} = -0.025\ \mathrm{C/m^2}$ and the numerically calculated value of the surface potential $\phi_\mathrm{s} = -33\,\mathrm{mV}$, a capacitance per unit area of $C_\mathrm{DL} = 77\ \mu\mathrm{F/cm^2}$ is obtained according to eq. (8.2). Very often, the double layer capacitance is instead approximated by the parallel plate capacity per unit area

$$C_{\mathrm{DL},\parallel} = \frac{\varepsilon_0 \varepsilon_\mathrm{r}}{\kappa^{-1}}. \tag{8.3}$$

Here, the Debye screening length κ^{-1} has the meaning of an (effective) 'thickness' of the electric double layer. In this case, using $\kappa^{-1} = 0.96\,\mathrm{nm}$ and $\varepsilon_\mathrm{r} = 78$, a capacity per unit area of $C_{\mathrm{DL},\parallel} = 72\ \mu\mathrm{F/cm^2}$ is obtained. In the latter case, the capacity does not depend on the surface charge density. It only depends on the salt concentration through the Debye screening length. Hence, the ability to store charge, i.e. the capacity, increases for higher salt concentrations.

The numerical Poisson–Boltzmann calculations for the capacitance (using the same data as in Fig. 8.4) are shown in Fig. 8.5. The capacitance increases rapidly for higher

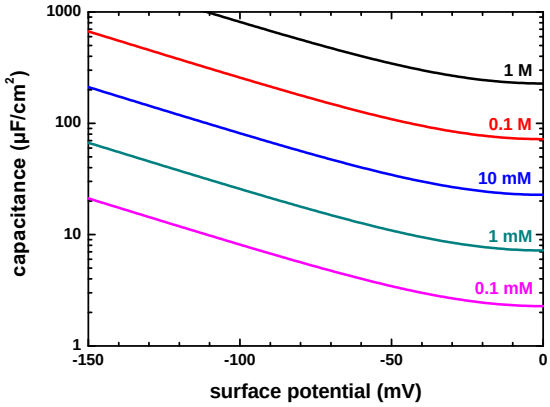


Figure 8.5.: Calculated capacitance of the electric double layer at different salt concentrations. The capacitance has a nearly constant value only for very small potentials.

potentials but at very small surface potentials, the capacitance is equal to the approximation of eq. (8.3). This is expected because in the limit of low potentials, the solution of the Poisson–Boltzmann equation must converge to the solution of the Debye–Hückel equation.

The following effects are not taken into account in this model: Ion-correlation, steric hindrance, finite ion size, ion adsorption, fixed charge distribution away from the interfaces (e.g. adsorbed molecules), hydrophobic or hydrophilic surfaces. Improvements of the Gouy–Chapman model very often discussed in literature are the Stern and Helmholtz layers. These models are useful in terms of giving a qualitative picture of the electric double layer and have been discussed in detail in Ref. [Lub06]. However, they are practically not very useful as in general no parameters for typical length scales, charge density accumulations and dielectric constants of specific arrangements are available introducing too many adjustable fitting parameters to the system. Therefore we chose to improve our electrolyte model by implementing a recently developed approach that uses potentials of mean force. Additionally, this model also allows for the description of hydrophilic and hydrophobic interfaces. This leads to a so-called extended Poisson–Boltzmann equation which is described in Section 8.4. Later, we will apply this model to graphene and diamond based biosensors and compare results with experiment.

8.2. Debye–Hückel approximation

The full Poisson–Boltzmann equation is a nonlinear differential equation for the electrostatic potential and describes long-ranged electrostatic interactions quite accurately. Very often, one is interested in reducing it to a simpler form which can be solved analytically. Within the Debye–Hückel (DH) approximation, the Poisson–Boltzmann equation

is linearized by expanding the exponential of eq. (7.3) up to first order in ϕ so that the potential distribution in the electrolyte is governed by

$$\left(\nabla^2 - \kappa^2\right)\phi(\mathbf{x}) = 0, \tag{8.4}$$

where the Debye screening length is given by

$$\kappa^{-1} = \left(\sum_{i=1}^{N} \frac{\varepsilon_\mathrm{r}\varepsilon_0 k_\mathrm{B} T}{(z_i e)^2 c_{i,0}}\right)^{1/2}. \tag{8.5}$$

The symbols have the same meaning as in eq. (7.3). The Debye screening length is often used as a descriptive parameter of the system of investigation and is of the order of a few nanometers. For instance, for an electrolyte with a planar surface at $x_0 = 0\,\mathrm{nm}$, the solution that satisfies the one-dimensional variant of eq. (8.4) is given by

$$\phi(x) = \phi_\mathrm{DH} \exp(-\kappa x), \tag{8.6}$$

where ϕ_DH is the Debye–Hückel potential at x_0. It is related to the (effective) surface charge density σ_DH at x_0 as follows (compare with eq. (8.2) and eq. (8.3))

$$\sigma_\mathrm{DH} = \frac{\varepsilon_\mathrm{r}\varepsilon_0 \phi_\mathrm{DH}}{\kappa^{-1}}. \tag{8.7}$$

One should keep in mind that the Debye–Hückel equation is only applicable for low electrostatic potentials where it holds

$$e\phi(\mathbf{x}) \ll k_\mathrm{B} T. \tag{8.8}$$

Effectively a diffuse double layer at low potential behaves like a parallel plate capacitor where the electrochemical double layer capacitance per unit area can be estimated by eq. (8.3). Therefore, κ^{-1} is often termed the 'thickness' of the double layer although this is somehow imprecise as the thickness is larger (compare with Fig. 8.3). For larger potentials the actual surface charge density σ_s or surface potential ϕ_s are substantially different from σ_DH and ϕ_DH (compare with Fig. 8.4). Typically, the Debye–Hückel approximation is used to estimate surface potentials and it is not used in numerical calculations. In Chapter 7 we compared the validity of the DH approximation to the Poisson–Boltzmann equation by solving both equations numerically for a situation where the electrolyte region contains a charged region of amino acids close to the surface.

8.3. Buffer solutions

In Chapter 7 we applied a Poisson–Boltzmann model to a silicon protein sensor where we have added a sophisticated model to our electrolyte that takes into account buffer ion concentrations as a function of pH value. To calculate the concentrations of the buffer ions, we briefly sketch the relevant equations that have been implemented into our self-consistent algorithm. While the details of buffer solutions have been described in detail in

Ref. [BE96] and are thus widely known, to our knowledge these equations have not been considered so far when modeling the pH dependence of solid–liquid systems. However, this is important as the buffer ion concentrations and thus the screening behavior of electrolytes, as well as chemical or biological reactions, depend strongly on the pH value. Therefore, it is useful to study the behavior of biosensors at various pH values, and at various salt concentrations using monovalent and divalent salts. Theoretical models that reproduce experimental results for a particular pH value can then be benchmarked if they will also work at other pH values and for other salt properties. If they do, this gives confidence into the model.

8.3.1. Ionic strength

The ionic strength of the electrolyte is defined as

$$I(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N} c_i(\mathbf{x}) z_i^2, \tag{8.9}$$

where N is the number of all different ion species that are present in the electrolyte, c_i is the concentration and z_i is the valency of the ion species i. Because the concentrations of the ions in the vicinity of the semiconductor surface depend on the spatial coordinates, our algorithm allows for a spatially varying ionic strength. In physiological systems the ionic strength is of the order 150 mM.

8.3.2. Effect of temperature on buffers

The parameter $\mathrm{d}\mathrm{p}K_\mathrm{a}/\mathrm{d}T$ defines the change in $\mathrm{p}K_\mathrm{a}$ with temperature. This quantity depends on the buffer, and can be negative or positive or even close to zero. Thus the temperature dependent $\mathrm{p}K_{\mathrm{a},T}$ value is given by

$$\mathrm{p}K_{\mathrm{a},T} = \mathrm{p}K_\mathrm{a} + \mathrm{d}\mathrm{p}K_\mathrm{a}/\mathrm{d}T \cdot (T - 298.15\,\mathrm{K})\,. \tag{8.10}$$

Here, T is given in units of Kelvin and the 'thermodynamic' $\mathrm{p}K_\mathrm{a}$ value is defined for 25 °C.

8.3.3. Debye–Hückel relationship

When using biological sensors, the pH is typically adjusted by titration and can be measured. Thus the pH of the bulk electrolyte is an input quantity for our simulations. Knowing the pH, one can calculate the concentrations of the buffer ions taking into account the temperature and the ionic strength of the solution. The $\mathrm{p}K'_{\mathrm{a},T}$ value determines the concentrations of the buffer ions but itself depends on the ionic strength $I(\mathbf{x})$ and on temperature T. As the ionic strength depends on the concentrations of the buffer ions, we have to solve this nonlinear equation self-consistently by an iterative scheme. The usually employed Debye–Hückel relationship reads

$$\mathrm{p}K'_{\mathrm{a},T} = \mathrm{p}K_{\mathrm{a},T} + (2z_a - 1)\left[\frac{A\sqrt{I}}{1+\sqrt{I}} - 0.1 \cdot I\right], \tag{8.11}$$

where $\mathrm{p}K'_{\mathrm{a},T}$ is called the 'modified' (or 'apparent' or 'working') $\mathrm{p}K_\mathrm{a}$ value, z_a is the charge on the conjugate acid species and the constant $A(T)$ depends on the temperature of the solution. The value of A is around 0.5 (at $T = 0$ °C: $A = 0.4918$, at $T = 100$ °C: $A = 0.6086$). $\mathrm{p}K'_{\mathrm{a},T}(\mathbf{x})$ is a function of position $\mathbf{x}$ because the ionic strength $I(\mathbf{x})$ is a function of position whereas $\mathrm{p}K_{\mathrm{a},T}$ only depends on the temperature.

8.3.4. Henderson–Hasselbalch equation

The Henderson–Hasselbalch equation

$$\mathrm{pH} = \mathrm{p}K'_{\mathrm{a},T} + \log_{10}\frac{[\mathrm{base}]}{[\mathrm{acid}]}, \tag{8.12}$$

relates the pH of the electrolyte to the $\mathrm{p}K'_{\mathrm{a},T}$ of the conjugate acid–base pair and the relative concentrations of acid and base. Since all quantities of this equation depend on spatial coordinates, the local pH value is also a function of position. In the vicinity of the semiconductor surface, the local pH therefore differs from the pH of the bulk electrolyte. Most buffers involve only one chemical reaction, thus a single $\mathrm{p}K_\mathrm{a}$ value is sufficient. Some buffers are more complicated and involve three reactions, e.g. the phosphate buffer saline (PBS) solution, which is used in Chapter 7, requires three $\mathrm{p}K_{\mathrm{a}i}$ values ($i = 1, 2, 3$). As the concentrations of the ions also depend on the electrostatic potential through the Poisson–Boltzmann equation (eq. (1.1) and eq. (7.3)) – which is influenced by the Schrödinger equation that determines the quantum mechanical charge density in the semiconductor device region – it is clear that only a numerical approach is feasible to solve this coupled system of equations self-consistently.

8.3.5. Phosphate buffer

Phosphate buffer saline (PBS) is made of orthophosphoric acid $\mathrm{H_3PO_4}$ and shows three dissociation reactions:

$$\mathrm{H_3PO_4} \overset{\mathrm{p}K'_{\mathrm{a}1,T}}{\rightleftharpoons} \mathrm{H_2PO_4^-} + \mathrm{H^+} \overset{\mathrm{p}K'_{\mathrm{a}2,T}}{\rightleftharpoons} \mathrm{HPO_4^{2-}} + 2\mathrm{H^+} \overset{\mathrm{p}K'_{\mathrm{a}3,T}}{\rightleftharpoons} \mathrm{PO_4^{3-}} + 3\mathrm{H^+} \tag{8.13}$$

Using the Henderson–Hasselbalch equation (eq. (8.12)), the concentrations of the involved ions can be calculated by the following formulas:

$$[\mathrm{H_3PO_4}] = \frac{[\mathrm{PBS}]}{1 + 10^{\mathrm{pH}-\mathrm{p}K'_{\mathrm{a}1,T}} \cdot \left(1 + 10^{\mathrm{pH}-\mathrm{p}K'_{\mathrm{a}2,T}} \cdot \left(1 + 10^{\mathrm{pH}-\mathrm{p}K'_{\mathrm{a}3,T}}\right)\right)} \tag{8.14}$$

$$[\mathrm{H_2PO_4^-}] = [\mathrm{H_3PO_4}] \cdot 10^{\mathrm{pH}-\mathrm{p}K'_{\mathrm{a}1,T}} \tag{8.15}$$

$$[\mathrm{HPO_4^{2-}}] = [\mathrm{H_2PO_4^-}] \cdot 10^{\mathrm{pH}-\mathrm{p}K'_{\mathrm{a}2,T}} \tag{8.16}$$

$$[\mathrm{PO_4^{3-}}] = [\mathrm{HPO_4^{2-}}] \cdot 10^{\mathrm{pH}-\mathrm{p}K'_{\mathrm{a}3,T}} \tag{8.17}$$

$$[\mathrm{Na^+}] = -z_1[\mathrm{H_2PO_4^-}] - z_2[\mathrm{HPO_4^{2-}}] - z_3[\mathrm{PO_4^{3-}}] \tag{8.18}$$

Here, $z_1 = -1$, $z_2 = -2$ and $z_3 = -3$ are the valencies of the respective ions $\mathrm{H_2PO_4^-}$, $\mathrm{HPO_4^{2-}}$ and $\mathrm{PO_4^{3-}}$. In our implementation, the concentration of the PBS buffer and the

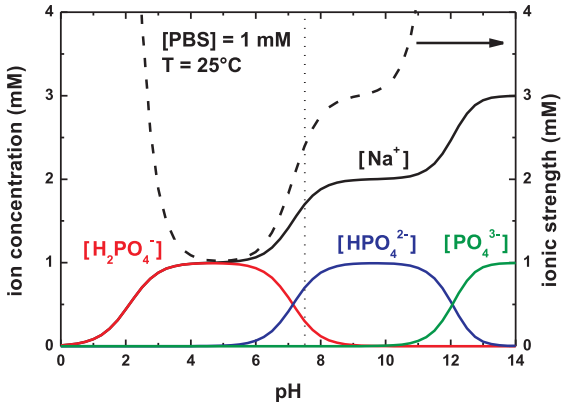


Figure 8.6.: Calculated concentrations of the buffer ions (solid lines) of the phosphate buffer saline (PBS) solution as a function of pH. At small and large pH values, the ionic strength (dashed line) strongly increases due to the increase in $[H_3O^+]$ and $[OH^-]$ concentrations, and their corresponding anions and cations. The influence of the valency on ionic strength (quadratic dependence, see eq. (8.9)) is noticeable for $[HPO_4^{2-}]$.

pH in the bulk electrolyte are fixed. However, the local value for the pH depends on the local concentration of H_3O^+ ions. The concentrations of the buffer ions $[H_2PO_4^-]$, $[HPO_4^{2-}]$, $[PO_4^{3-}]$ and $[Na^+]$ are then calculated using the parameters listed in Table 8.1. For a given local value of pH, equations (8.9), (8.11) and (8.12) (i.e. equations (8.14), (8.15), (8.16), (8.17) and (8.18)) have to be solved self-consistently in an iterative manner. Figure 8.6 shows the concentrations of the buffer ions and the ionic strength as a function of pH for a 1 mM PBS buffer. The second column (0 mM KCl) of Table 7.1 lists these values at pH = 7.5 (vertical dotted line in Fig. 8.6).

Table 8.1.: Buffer parameters: Phosphate buffer saline (PBS)

Symbol	Value	Units
$pK_{a1}(25\ °C)$	2.15	
$pK_{a2}(25\ °C)$	7.21	
$pK_{a3}(25\ °C)$	12.33	
dpK_{a1}/dT	0.0044	K^{-1}
dpK_{a2}/dT	−0.0028	K^{-1}
dpK_{a3}/dT	−0.026	K^{-1}
z_{a1}	0	
z_{a2}	−1	
z_{a3}	−2	
$A(25\ °C)$	0.5114	

8.4. Extended Poisson–Boltzmann equation: Potentials of Mean Force

Various corrections to the Poisson–Boltzmann equation have been proposed and are reviewed in Ref. [BKN+05]. In this section we describe the extended Poisson–Boltzmann (ePB) equation that includes corrections through a potential energy term. It treats the long-ranged electrostatic interactions between ions and the net surface charge and among the ions at finite bulk salt concentrations $c_{i,0}$ on an approximate mean-field level and simultaneously takes into account ion-specific surface interactions through potentials of mean force (PMF). The PMFs are defined for each ion depending on a hydrophobic (nonpolar) or hydrophilic (polar) solid–electrolyte interface. This approach is based on the work of Schwierz, Horinek and Netz [SHN10] and allows us to distinguish between the behavior of different ions, like that the repulsion of F^- ions is much stronger than for Cl^- or I^- ions at hydrophobic surfaces. They published fitting functions for the PMFs obtained from molecular dynamics simulations that we will use throughout this thesis. In fact, these functions have been obtained from hydrophobic CH_3-terminated and from hydrophilic OH^--terminated self-assembled monolayers and are valid for planar surfaces. In the following chapters we apply this model to diamond and graphene based sensor structures. Obviously, our interfaces are different compared to self-assembled monolayers. Thus we consider the provided fitting functions as a model system in order to be able to compare hydrophobic vs. hydrophilic semiconductor–electrolyte interfaces. The ion density in the electrolyte is now given by

$$\rho(x) = \sum_{i=1}^{N} z_i e c_{i,0} \exp\left(-\frac{z_i e\left(\phi\left(x\right) - U_\mathrm{G}\right) + V_{\mathrm{PMF},i}(x)}{k_\mathrm{B} T}\right), \tag{8.19}$$

where the potential energy term $V_{\mathrm{PMF},i}(x)$ is the spatially varying potential of mean force of ion species i. The other symbols have the same meaning as in eq. (7.3). The PMFs are only relevant in the vicinity of the interface (0 nm to 1.4 nm) and introduce a repulsive term which prevents the unphysical situation of too many ions coming too close to the surface.

Instead of assuming a constant value of $\varepsilon_\mathrm{r}^{\mathrm{H_2O}} = 78$ for the dielectric constant of water, for all PMF calculations a local dielectric constant for the electrolyte based on the water density will be employed, hence including effects such as the water replacement by the distribution of ions at surfaces. We assume the local dielectric constant $\varepsilon_\mathrm{r}(x)$ that enters the Poisson–Boltzmann equation to vary in the electrolyte as a function of distance x from the solid–liquid interface at $x_0 = 0\,\mathrm{nm}$ like

$$\varepsilon_\mathrm{r}(x) = \varepsilon_{\mathrm{r,s}} + \left(\varepsilon_\mathrm{r}^{\mathrm{H_2O}} - \varepsilon_{\mathrm{r,s}}\right) \frac{\rho^{\mathrm{H_2O}}(x)}{\rho_0^{\mathrm{H_2O}}}. \tag{8.20}$$

Here, the dielectric constant is assumed to be proportional to the water density profile $\rho^{\mathrm{H_2O}}(x)$ (see Fig. 8.7) where $\rho_0^{\mathrm{H_2O}}$ is the bulk density of water. This density profile has been obtained by a fitting function. The fitting parameters were chosen to match

the according profiles obtained from molecular dynamics simulations (see [SHN10]). In fact, this density profile is different for hydrophobic compared to hydrophilic surfaces as shown in Fig. 8.8. Thus the microscopic structuring of water in the vicinity of a nonpolar or a polar solid wall is taken into account in our model. $\varepsilon_{\mathrm{r,s}}$ is the relative dielectric constant at the surface of the electrolyte which can be the constant of e.g. a self-assembled monolayer or the constant of vacuum ($\varepsilon_{\mathrm{r,s}} = 1$) if there is a distance of a few Angstrom where there are no ions or water molecules at the solid–electrolyte interface. For the calculations in this section, we assume $\varepsilon_{\mathrm{r,s}} = 4$ as in Ref. [SHN10] for a self-assembled monolayer but in the following sections on graphene and diamond

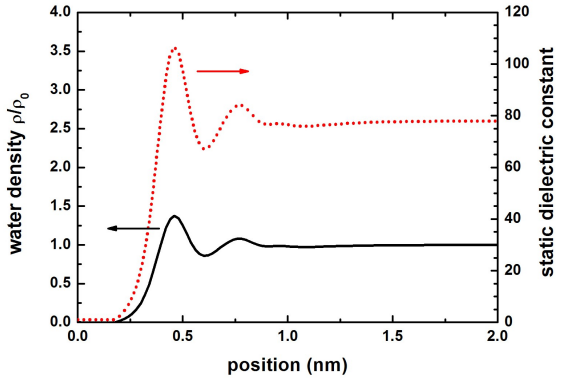


Figure 8.7.: Spatially varying static dielectric constant $\varepsilon_{\mathrm{r}}(x)$ (red dotted line) of the electrolyte at a hydrophobic solid–liquid interface according to the parameters of Ref. [SHN10]. The static dielectric constant varies from $\varepsilon_{\mathrm{r}} = 1$ at the interface to $\varepsilon_{\mathrm{r}} = 78$ in the bulk electrolyte. It is proportional to the water density profile (black solid line).

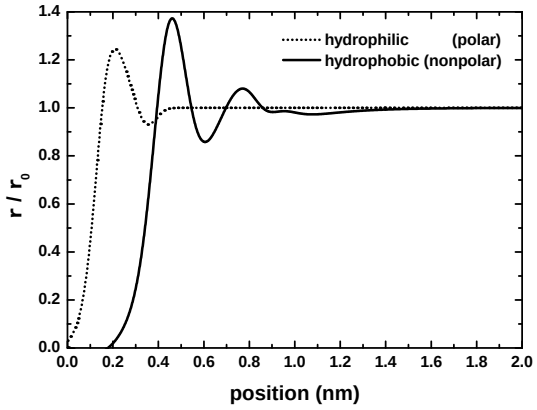


Figure 8.8.: Resulting water density profile (fitting function) at hydrophobic (solid line) and hydrophilic (dotted line) solid–liquid interfaces.

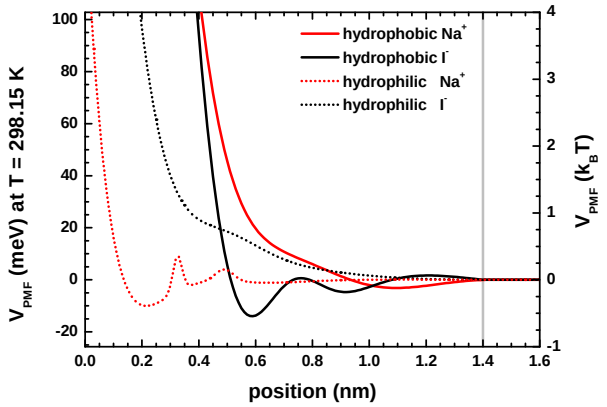


Figure 8.9.: Ionic potentials of mean force (PMF) for the ions Na^+ (black lines) and I^- (red lines) at a hydrophobic (solid lines) and a hydrophilic solid–liquid interface (dotted lines). The interface is at 0 nm, and the PMFs are zero beyond 1.4 nm indicated by the vertical line.

based sensors we use $\varepsilon_{r,s} = 1$. The water density only varies along the first 2 nm away from the interface. The water is depleted from the interface in the hydrophobic case by approximately 0.3 nm. In general, the water density profile will differ slightly for different solids, e.g. it was found that for diamond and graphene the spatial variation of the density had roughly the same shape but the maxima and minima in graphene were more pronounced leading to a slightly different dielectric constant profile. The reason for this is probably that the first atomic layer (which has by far the biggest influence) of graphene is denser than the first layer of diamond [Bon]. For the purpose of our model where we are mainly interested in qualitative trends, these deviations are thus negligible from our current point of view. However, as we show in Fig. 10.14, a shift of the water density profile of ±0.05 nm, i.e. a shift in the onset, changes the results slightly. In contrast, we found that shifting the PMFs by this amount has only negligible influence. In fact, the PMFs are quite robust, for instance calculations on various hydrophobic surfaces showed nice agreement [Net]. This gives us confidence in applying the PMFs, derived for the interface of self-assembled monolayers in contact to water, to our graphene and diamond based solid–liquid interfaces in the next two chapters.

As a simple example to illustrate the extended Poisson–Boltzmann equation, we model a 50 nm electrolyte solution containing 1 M or 10 mM of NaI. It is the example used by Schwierz et al., where they compared their predictions of ion distributions at various salt concentration successfully against molecular dynamics simulations, demonstrating the robustness of the extended PB equation. We assume solid interfaces at the left and right boundary of the electrolyte that are either both hydrophobic or hydrophilic.

Figure 8.9 shows the PMF of Na^+ and I^- ions at hydrophobic and hydrophilic surfaces as a function of distance from the interface (see Supporting Information of Ref. [SHN10]

for more detailed information). Very large values for the potential energies indicate strong repulsion of the ions, implying no ions close to the surface. The repulsion of ions at hydrophobic interfaces is obviously stronger. Further away from the interface the repulsion is modulated by the water density, and there is even a region of attraction for I^- at around 0.6 nm for the hydrophobic interface. However, we note that the actual shape of the potentials of mean force does not have much influence on our calculations in Chapter 9 and Chapter 10 because we are interested in the carrier density in the solid, and not so much in the actual ion distribution in the electrolyte. What has significant influence is the onset of the strong repulsion of the ions, and the dielectric constant of water. As these two ingredients strongly differ for hydrophobic and hydrophilic solids, significantly different results are obtained for these two cases. Additionally, when completely ignoring PMFs as in the standard Poisson–Boltzmann model (PB), the results will be even more different. In any case, the hydrophilic results are expected to lie in between the results of the hydrophobic and the PB model. In this thesis we are interested in showing the differences among the three models: hydrophobic solid, hydrophilic solid, standard Poisson–Boltzmann model. In most cases our interfaces are charged (e.g. the charge in a two-dimensional electron gas in the solid), thus always the ion type with the opposite charge strongly dominates the ion concentration profiles at the interface and its concentration is much higher than its equilibrium concentration. Consequently, we do not care much about the tiny modulations in the PMF profiles. They are however noticeable for uncharged solid interfaces. In this case the ion concentration profiles oscillate around their equilibrium value to a lesser extent and thus strongly follow the shape of their PMFs. This can be seen for the hydrophobic case in Fig. 8.10 and for the hydrophilic case in Fig. 8.11, respectively, where the concentrations of the Na^+ ions are shown in red, and the I^- ions in black. The vertical lines indicate the barrier from where on the PMFs are zero and the usual Poisson–Boltzmann screening behavior takes place. The concentration of I^- ions is stronger at the hydrophobic surface than for Na^+ ions indicating that their adsorption is stronger. This situation is reversed at hydrophilic surfaces. In both cases the preference of one ion over the other decreases with increasing ionic strength. Thus the shapes of the ion density profiles $c_i/c_{i,0}$ for 1 M (solid lines) and 10 mM solutions (dotted lines) are not identical, although within our model, the water density and the PMFs are independent of the equilibrium concentration of the salt in the solution. A self-consistent solution of the extended Poisson–Boltzmann equation not only takes into account the PMFs and the water density but also takes care of the local electrostatic potential and the spatial variation of the ionic strength (eq. (8.9)) that dominates the complicated overall screening behavior of the electrolyte at the interface. The results of our calculations show perfect agreement to Fig. 3(A) and Fig. 3(C) of Ref. [SHN10].

Figure 8.12 shows the electrostatic potential $-\phi(x)$ for the hydrophobic surface for 1 M (solid lines) and 10 mM (dotted lines) NaI solutions. The 1 M solution has a higher ionic strength and thus interface effects vanish after approximately 2 nm completely, whereas for the 10 mM solution the screening is not so effective. The red lines show the same data on a logarithmic scale (arrows) which are in perfect agreement to Fig. 4(A) of Ref. [SHN10].

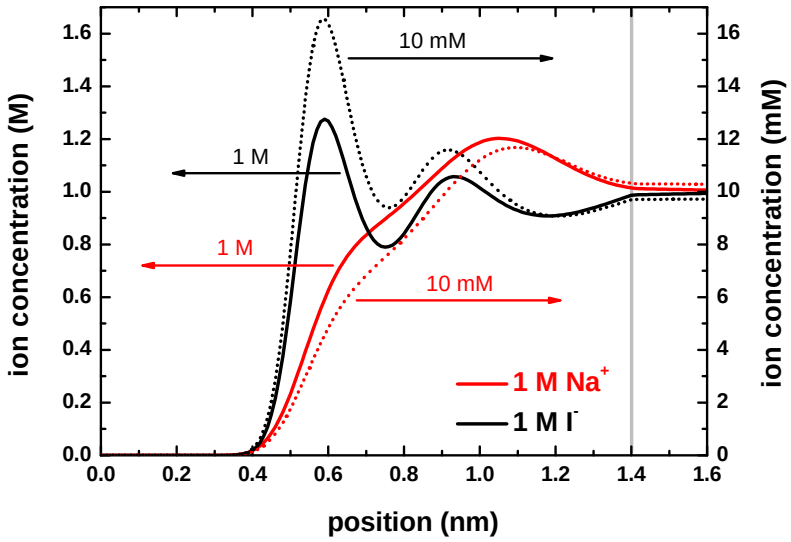


Figure 8.10.: Ion concentration profiles of 1 M (solid lines) and 10 mM (dotted lines) NaI solution at a hydrophobic (nonpolar) surface (Na^+ ions: red lines, I^- ions black lines)

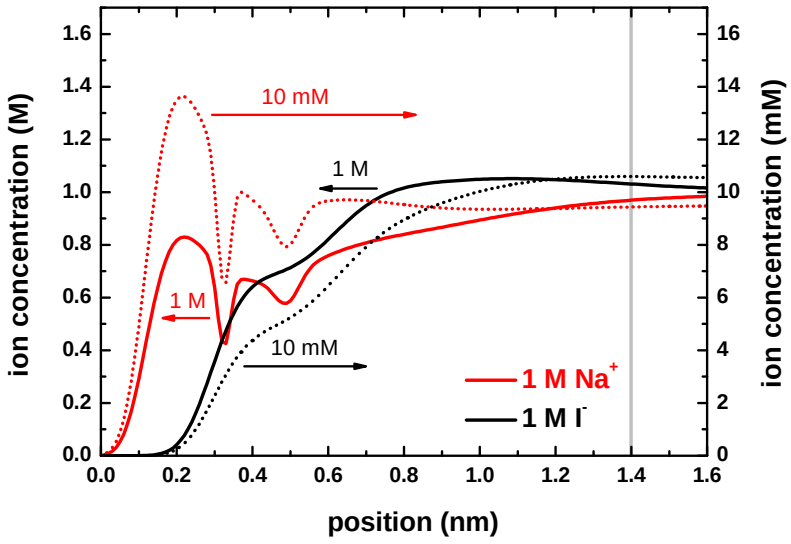


Figure 8.11.: Ion concentration profiles of 1 M (solid lines) and 10 mM (dotted lines) NaI solution at a hydrophilic (polar) surface (Na^+ ions: red lines, I^- ions black lines)

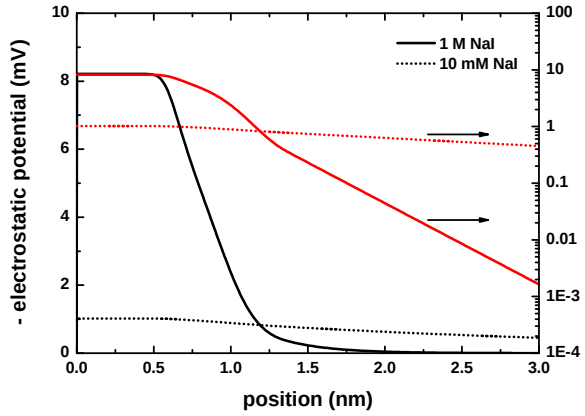


Figure 8.12.: Electrostatic potential of 1 M (solid lines) and 10 mM NaI (dotted lines) solution at a hydrophobic surface. The red lines show the same data on a logarithmic scale (arrows).

In the following two sections, our electrolytes contain Na^+ and Cl^- ions, therefore we show their potentials of mean force (PMF) for the hydrophobic (solid lines) and hydrophilic interfaces (dotted lines) in Fig. 8.13. We note that the PMF for Na^+ ions is identical to the one of Fig. 8.9. The PMFs are zero beyond 1.4 nm indicated by the vertical line. One can clearly see that the PMFs for the hydrophobic interface repel the ions strongly for distances below 0.4 nm from the interface.

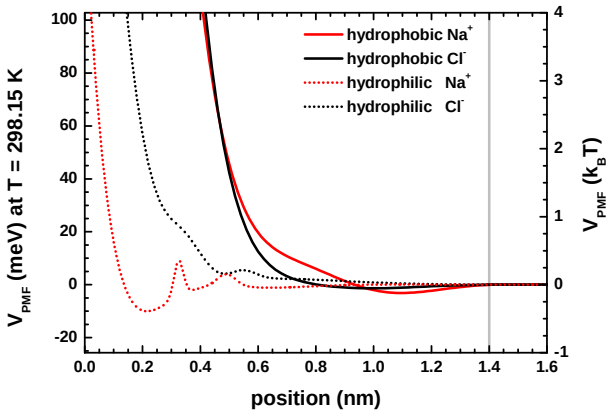


Figure 8.13.: Ionic potentials of mean force (PMF) for the ions Na^+ (black lines) and Cl^- (red lines) at a hydrophobic (solid lines) and a hydrophilic solid–liquid interface (dotted lines). The PMFs repel the ions strongly for distances below 0.4 nm from the interface for the hydrophobic case.

9. Modeling graphene based solution gated field-effect transistors

In this chapter we model graphene based solution gated field-effect transistors (SGFET). First, we discuss the band structure of graphene. Then we compare the density and capacitance of ideal graphene layers with graphene layers that are subject to potential fluctuations. Finally, we model graphene based sensors in liquid environments. We apply our new extended Poisson–Boltzmann approach (see Section 8.4) and compare its results to simpler models and to experiment. The effect of the solution-gate potential on the electronic properties of graphene is explained using a model that takes into account the microscopic structure of water at the graphene–electrolyte interface.

9.1. Band structure of graphene

Graphene is a two-dimensional crystal which consists of a monolayer of graphite. Although its band structure and related properties have been studied and known since decades [Wal74], only recently the material has been subject of intensive research worldwide mainly due to its exceptional physical properties and numerous potential applications in electronics but also due to its low-cost fabrication techniques (e.g. 'Scotch tape technique'). Eventually, in 2010 the Nobel prize was awarded to A. Geim and K. Novoselov for their pioneering work [NGM$^+$04]. Compared to silicon, graphene shows superior chemical stability and is expected to be bio-inert which makes it an ideal material for biosensor and bioelectronic applications. Its electronic properties allow it to outperform silicon for sensing devices because graphene is an ideal two-dimensional system with very high mobilities for both electrons and holes even at room temperature. In addition, it is sensitive to environmental conditions and charge adsorption. So far most of the reports on graphene field-effect transistors have addressed operation under vacuum or atmospheric conditions. Recently also operation under aqueous electrolyte environments were demonstrated for pH sensing and protein adsorption [ACWL08, HCM$^+$09]. A summary of the technological challenges and references to recent work can be found in Ref. [DHL$^+$10]. However, a detailed understanding of the graphene–electrolyte interface and the effect of the electrolyte on the electronic transport in graphene is still lacking. The former will be addressed in this thesis.

In this section we analyze the band structure of graphene using a tight-binding approach in the nearest-neighbor (nn) and third-nearest-neighbor (3^{rd}-nn) approximation. We compare three different sets of parameters with the widely used linear $E(\mathbf{k})$ relationship.

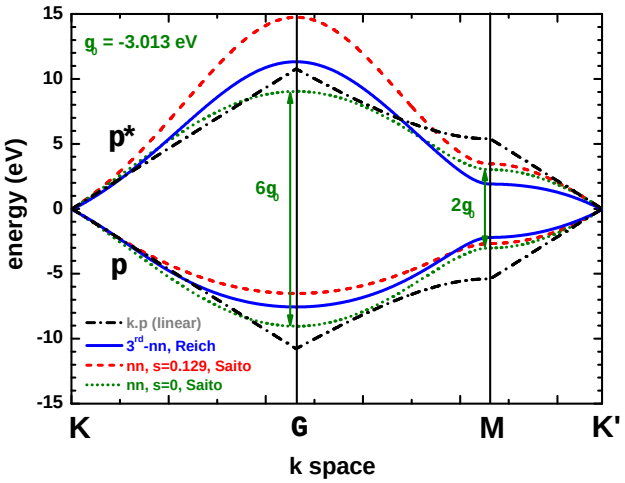


Figure 9.1.: Calculated band structure of graphene using the tight-binding method. Conduction band π^* (upper part) and valence band π (lower part) of graphene along special high symmetry directions in the two-dimensional hexagonal Brillouin zone for different models.

The conduction (E_+) and valence band energies (E_-) can be calculated by

$$E_\pm = \frac{E_{2p} \mp \gamma_0 w(\mathbf{k})}{1 \mp s_0 w(\mathbf{k})}, \tag{9.1}$$

where E_{2p} is the site energy (orbital energy) of the $2p_z$ atomic orbital, and γ_0 is the C–C transfer energy which is typically in the interval $-3\,\mathrm{eV} < \gamma_0 < -2.5\,\mathrm{eV}$. The overlap of the electronic wave function on adjacent sites is denoted with s_0. Since it is a small value, it is often neglected. Finally, $w(k)$ is given by [SK01]

$$w(\mathbf{k}) = \sqrt{1 + 4\cos\frac{ak_y}{2}\cos\frac{\sqrt{3}ak_x}{2} + 4\cos^2\frac{ak_y}{2}}. \tag{9.2}$$

Figure 9.1 shows the conduction band π^* (E_+) and valence band π (E_-) of graphene along special high symmetry directions in the two-dimensional hexagonal Brillouin zone in $\mathbf{k}$ space. The high symmetry points that are used in this graph (from left to right) are: $\mathrm{K} = \left(0, \frac{2}{3}\right)\frac{2\pi}{a}$, $\Gamma = (0,0)$, $\mathrm{M} = \left(\sqrt{\frac{1}{3}}, 0\right)\frac{2\pi}{a}$, $\mathrm{K}' = \left(\sqrt{\frac{1}{3}}, \frac{1}{3}\right)\frac{2\pi}{a}$. The blue solid lines are the third-nearest-neighbor tight-binding approximation using the parameters of Reich et al. [RMTO02]. Using this set of parameters, the band gap at the K and K′ points is not exactly zero. The red dashed lines are the nearest-neighbor approximation using the parameters of Saito [SK01]. They are nonsymmetric with respect to the electron and hole dispersion and are close to calculations from first principles and experimental

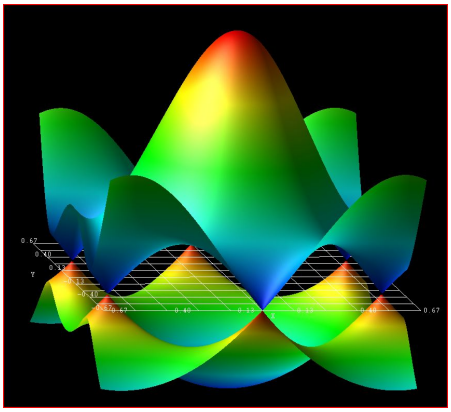


Figure 9.2.: Calculated energy dispersion $E(k_x, k_y)$ of graphene using the parameters of Saito with $s_0 = 0.129$ in the nearest-neighbor approximation. The upper part refers to the conduction band, the lower part to the valence band. At the six Dirac points where the conduction band edge energy equals the valence band edge energy, the dispersion becomes linear.

data. Setting $s_0 = 0$ yields the green dotted lines that are symmetric with respect to the Fermi level $E_\mathrm{F} = 0\,\mathrm{eV}$. Then the dispersion of both π^* and π is the same (apart from the sign). In this case the splitting energy at Γ is three times as large as at the M point (indicated by the arrows). The black dash-dotted lines correspond to the linear $E(k)$ relationship that is typically used ($\mathbf{k} \cdot \mathbf{p}$ approximation or linear expansion). It is valid at the K and K$'$ points for low energies, and this is the approximation that we will use in the following sections. The linear dispersion is independent of the parameter s_0. Thus for small values of $\mathbf{k}$ (with respect to the K point), the energy dispersion can be approximated by a linear dispersion relation

$$E(\mathbf{k}) = E_{2p} \pm \hbar v_\mathrm{F} \left|\mathbf{k}\right| = E_{2p} \pm \sqrt{3}\gamma_0 \frac{a}{2} k, \tag{9.3}$$

where a is the lattice constant of graphene ($a = 0.24612\,\mathrm{nm}$), $k = \left(k_x^2 + k_y^2\right)^{1/2}$ and the Fermi velocity of the charge carriers is given by $v_\mathrm{F} = \sqrt{3}\left|\gamma_0\right| \frac{a}{2\hbar} \cong 0.98 \cdot 10^6\,\mathrm{m/s} \cong 0.003c$, where c is the velocity of light. At the K and K$'$ points, the band gap is zero.

Figure 9.2 shows the calculated energy dispersion $E(k_x, k_y)$ of graphene using the parameters of Saito with $s_0 = 0.129$ in the nearest-neighbor approximation. The upper part refers to the conduction band, the lower part to the valence band. They are not symmetric for the parameterization of Saito (see Fig. 9.1). At the six Dirac points (three K and three K$'$ points) where the conduction band edge energy equals the valence band edge energy, and thus the band gap is zero, the dispersion becomes linear. The point in the middle is the Γ point. The x axis shows k_x within the interval $[-2/3, 2/3]\,\frac{2\pi}{a}$, the same holds for k_y.

9.2. Density and capacitance of graphene films

In this section, we calculate the density in graphene layers according to Ref. [FKXJ07] assuming a linear energy dispersion $E(\mathbf{k})$ leading to symmetric electron densities n and hole densities p with respect to $|\eta|$. They are calculated to be

$$n = \frac{2}{\pi}\left(\frac{k_\mathrm{B}T}{\hbar v_\mathrm{F}}\right)^2 F_1\left(+\eta\right) \tag{9.4}$$

$$p = \frac{2}{\pi}\left(\frac{k_\mathrm{B}T}{\hbar v_\mathrm{F}}\right)^2 F_1\left(-\eta\right), \tag{9.5}$$

where the Fermi velocity v_F of the charge carriers in graphene was chosen to be $v_\mathrm{F} = 0.98 \cdot 10^6\,\mathrm{m/s}$. F_1 is the Fermi–Dirac integral of the order 1 having $\eta = \left(E_\mathrm{F} - E_\mathrm{D}\right)/k_\mathrm{B}T$ as its argument, where the position of the Fermi level E_F relative to the Dirac point E_D determines the charge density. Usually E_D is assumed to be at $E_{\mathrm{D},0} = 0\,\mathrm{eV}$ due to the symmetry of the energy dispersion. Fermi–Dirac integrals of any order can be evaluated numerically very efficiently using approximation formulas [Ant93]. For reasons that become clear in the next section, we assume the Fermi level to be constant and fixed at $E_\mathrm{F} = 0\,\mathrm{eV}$, i.e. the position of the Dirac point equals 0 eV only in the case of zero electrostatic potential

$$E_\mathrm{D} = E_{\mathrm{D},0} - e\phi(x). \tag{9.6}$$

Figure 9.3 shows the calculated electron and hole sheet carrier densities as a function of the position of the Fermi level at room temperature. Usually one is interested in small Fermi level variations. Figure 9.4 shows the same data in the regime of lower applied voltages (zoom of Fig. 9.3).

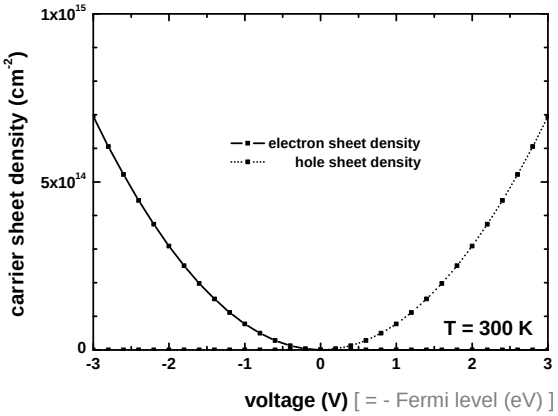


Figure 9.3.: Calculated electron and hole sheet densities in graphene as a function of position of the Fermi level for $T = 300\,\mathrm{K}$.

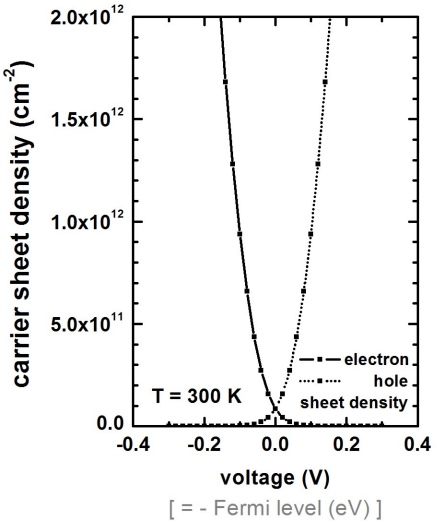


Figure 9.4.: Calculated electron and hole sheet densities in graphene as a function of position of the Fermi level for $T = 300\,\mathrm{K}$ in the low voltage regime.

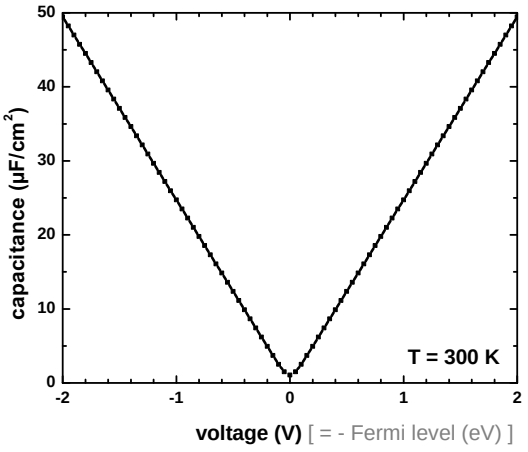


Figure 9.5.: Quantum capacitance of an ideal graphene layer at room temperature.

The quantum capacitance, $C = \partial Q/\partial U$, of ideal graphene is shown in Fig. 9.5. It is obtained by calculating the derivative of the total charge, $Q = n + p$, with respect to the voltage (i.e. local channel electrostatic potential).

An improvement to this ideal quantum capacitance model has been proposed by Xu et al. [XZP11]. They assumed that a real graphene layer differs from a perfect graphene layer due to potential fluctuations that obey a Gaussian distribution. The potential fluctuations lead to a fluctuation of the local density of states. They calculated the

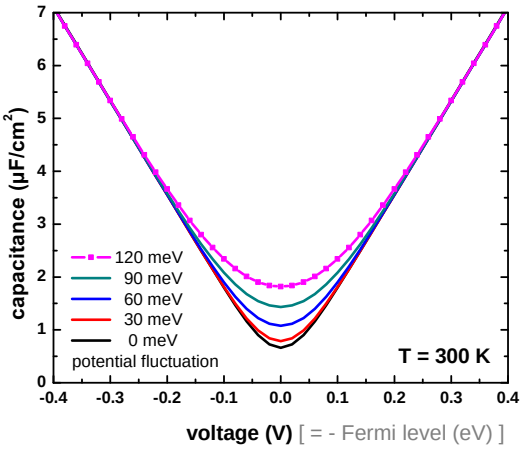


Figure 9.6.: Quantum capacitance of a nonideal graphene layer at room temperature for different values of the potential fluctuations.

average quantum capacitance of a large graphene sheet with fluctuating local potential by a microscopic capacitance model in which the macroscopic graphene plane is divided into N small cells with the same area. Essentially, they view the graphene plane as composed of many cells connected in parallel. We implemented their model slightly differently as our numerical Schrödinger–Poisson solver needs an expression for the density rather than the capacitance. The density in graphene including potential energy fluctuations δE is thus given by

$$n'(E_{\mathrm{FD}}, \delta E) = \int_{-\infty}^{\infty} n(E) P(E, E_{\mathrm{FD}}, \delta E) \mathrm{d}E, \tag{9.7}$$

where $E_{\mathrm{FD}} = E_{\mathrm{F}} - E_{\mathrm{D}}$ is the difference in energy of the Fermi level with respect to the Dirac point, and $n(E)$ is the equation for the density in ideal graphene (eq. (9.4) and eq. (9.5)). The potential energy fluctuations are assumed to follow a Gaussian distribution given by

$$P(E, E_{\mathrm{FD}}, \delta E) = \frac{1}{\sqrt{2\pi}\delta E} \exp\left(-\frac{(E - E_{\mathrm{FD}})^2}{2\delta E^2}\right). \tag{9.8}$$

Using this model, the quantum capacitance with respect to the voltage is shown in Fig. 9.6 for five different values of the potential fluctuation. In order to compare our results with the calculations of Xu et al., we used a Fermi velocity of $1.15 \cdot 10^6$ m/s in this figure. Only the quantum capacitance at low voltages is affected by potential fluctuations where they are responsible for the lower limit of the quantum capacitance. For large voltages the potential fluctuations are negligible because the density is already very high.

Finally, Fig. 9.7 shows the calculated intrinsic carrier density in thermal equilibrium and under no external perturbation ($V = 0$) where the Fermi level is at the Dirac point,

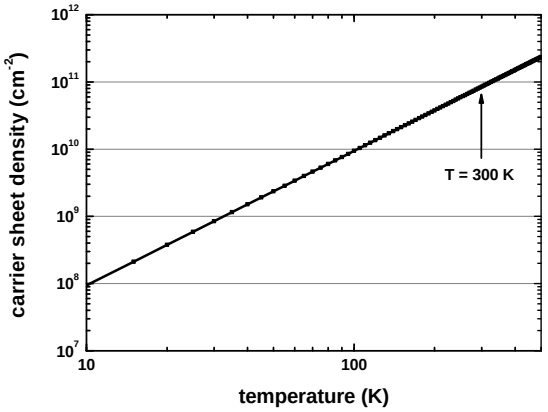


Figure 9.7.: Calculated intrinsic sheet carrier density in graphene vs. temperature. The arrow indicates the intrinsic carrier density at room temperature.

i.e. exactly in the middle of the (zero) band gap energy. The intrinsic carrier density depends quadratically on the temperature. At room temperature ($T = 300\,\mathrm{K}$) it is about $n_i = p_i = 8.5 \cdot 10^{10}\,\mathrm{cm}^{-2}$.

9.3. Results: Modeling graphene based sensors in liquid environments

In this section we describe our approach for modeling graphene based sensors in liquid environments and present our results. First we discuss the operation principle of a graphene solution-gated field-effect transistor. Figure 9.8 shows the effective modulation of the graphene conductivity by the electrolyte potential. Using the concept of an ideally polarizable graphene–electrolyte interface, the modulation of the conductivity by the electrolyte potential can be described as follows: The reference electrode is used to control the potential at the graphene–electrolyte interface. An applied gate potential fixes the potential drop between the Fermi level E_F in graphene and the reference electrode. From now on, the electrolyte potential will be referred to as gate potential U_G. For gate potentials more negative than the Dirac point (left part of figure), the position of E_F is driven further below the valence band maximum, increasing the number of holes in the graphene film and thus the conductivity. If the potential is reversed and more positive gate potentials are applied (right part of figure), E_F in the graphene film will be pushed above the conduction band minimum, increasing the number of electrons in graphene. This symmetric switching between electrons and holes is due to the semimetallic nature of graphene (see Fig. 9.3, Fig. 9.4). Therefore, the applied gate voltage modulates both the type and the number of free carriers. However, in order to fully understand the modulation of the carrier density in graphene by the electrolyte potential, it is necessary

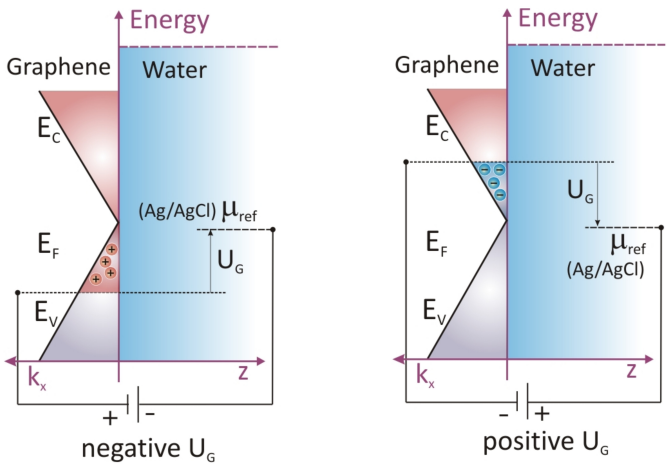


Figure 9.8.: Operation principle of a graphene solution-gated field-effect transistor. The schematic drawing demonstrates the modulation of the carrier density in the graphene layer. The applied gate voltage U_G with respect to the reference electrode shifts the Fermi level E_F in graphene below (left) or above (right) the Dirac point, thus modulating both the type and the number of free carriers. A negative value of U_G leads to an access of holes (left) whereas a positive U_G induces an electron gas (right).

to examine the charge distribution at the graphene–electrolyte interface. An electric double layer is expected to form at an ideally polarizable electrode–electrolyte interface. For carbon based electrodes this interfacial layer is typically described by a double layer capacitance varying from a few $\mu\mathrm{F/cm^2}$ for diamond electrodes to a few tens of $\mu\mathrm{F/cm^2}$ for graphite electrodes. For graphene not only the double layer capacitance is relevant. The quantum capacitance of the graphene film (Fig. 9.5, Fig. 9.6) must be taken into account as well.

A much simpler approach than ours for modeling liquid-gated graphene field-effect transistors has been presented by Heller et al. [HCM+09]. They compared liquid gating vs. back gating and found a striking difference in gating efficiency, namely that the liquid gating is more efficient (strong coupling). They conclude that the strength of the gate coupling determines the induced potential shifts that tune the Dirac point relative to the Fermi level, which in turn determines the number of electrons and holes in the conducting channel. They correctly point out that the total capacitance C_total (interfacial capacitance) of liquid gating is a series capacitance of the liquid gate capacitance (electric double layer capacitance) C_DL and the quantum capacitance C_Q of graphene

$$\frac{1}{C_\mathrm{total}} = \frac{1}{C_\mathrm{DL}} + \frac{1}{C_\mathrm{Q}}. \tag{9.9}$$

Consequently, the applied gate potential drops over these two capacitances and the applied gate voltage cannot be directly assigned to the electrostatic potential in graphene since part of the voltage drops in the electrolyte close to the surface. In the case of back

gating (weak coupling) where an insulating layer (SiO_2) is below the graphene in Heller's sample, the geometric capacitance of the oxide is in series to the quantum capacitance. Heller et al. calculate the total capacitance by assuming a constant value for the liquid gate capacitance based on a parallel plate capacitor model, although they mention that it is in fact gate dependent. The parallel plate capacitance per unit area is given by eq. (8.3). They assume $\kappa^{-1} = 1\,\mathrm{nm}$, corresponding to the Debye screening length of 0.1 M monovalent salt (compare with Fig. 8.2), and $\varepsilon_\mathrm{r} = 80$ for water. Thus they derive a (constant) value for the liquid gate capacitance of 70 $\mu\mathrm{F/cm^2}$. In contrast, our model allows us to take into account the gate dependent liquid gate capacitance, as well as a spatially varying dielectric constant ε_r for water. We calculated self-consistently the spatial charge $\rho(x)$ and electrostatic potential $\phi(x)$ distribution in the SiC–graphene–electrolyte system by solving the nonlinear Poisson equation (eq. (1.1)). We use the Dirichlet boundary condition $\phi(\infty) = U_\mathrm{G}$ for the electrostatic potential in the bulk electrolyte which is determined by the voltage U_G of the reference electrode (corresponding to zero net ion charge density in the electrolyte far away from the interface), and the Neumann boundary condition $\frac{\partial\phi}{\partial x} = 0\ \mathrm{V/m}$ (vanishing electric field) deep in the SiC layer corresponding to overall charge neutrality. The sheet charge density in graphene is calculated as described in Section 9.2. The Fermi level is assumed to be constant and fixed at $E_\mathrm{F} = 0\,\mathrm{eV}$, i.e. the energetic position of the Dirac point E_D equals 0 eV only in the case of zero electrostatic potential (eq. (9.6)). Consequently, an applied gate voltage U_G in the electrolyte modifies the electrostatic potential in the graphene layer and thus alters its charge density by moving the Dirac point with respect to the Fermi level. For the graphene layer we assume a thickness of 0.334 nm, corresponding to half the lattice spacing in graphite, i.e. the thickness of one monolayer. The dielectric constants of SiC and graphene were chosen to be $\varepsilon_\mathrm{r} = 10.3$ and $\varepsilon_\mathrm{r} = 5.68$, respectively, where the latter is actually the one for diamond. The dielectric constant in the electrolyte is proportional to the water density according to Ref. [SHN10] (see Fig. 8.7) and varies from $\varepsilon_\mathrm{r} = 1$ at the interface to $\varepsilon_\mathrm{r} = 78$ further away from the interface. The distribution of the ions in the electrolyte is calculated using an extended Poisson–Boltzmann approach that takes into account recently published [SHN10] ionic potentials of mean force (PMFs) $V_{\mathrm{PMF},i}(x)$ ($i = \left\{\mathrm{Na^+}, \mathrm{Cl^-}\right\}$). They were described in detail in Section 8.4. The ion density is given by eq. (8.19). The temperature has been assumed to be room temperature ($T = 298.15\,\mathrm{K}$). The Poisson equation has been discretized on a nonuniform grid using the finite differences method. It is solved numerically with a Newton–Raphson scheme. As CPU time is not critical (order of seconds) the grid spacing has been chosen to be very small (0.02 nm) to resolve the fitting functions of the potentials of mean force reasonably well at the solid–liquid interface. More details on the simulation of the combined system of semiconductor–electrolyte systems with the next**nano**^{3} software are described in Chapter 7. The PMFs describe hydrophobic (i.e. nonpolar) or hydrophilic (i.e. polar) solid–liquid interfaces and are based on fitting functions obtained from molecular dynamics simulations [SHN10]. They are shown in Fig. 8.13 together with the ones for hydrophilic interfaces. The PMFs have the effect of repelling the ions from the interface and are zero at distances larger than 1.4 nm from the interface. We compare this approach with the traditional Poisson–Boltzmann (PB) equation where no PMFs are

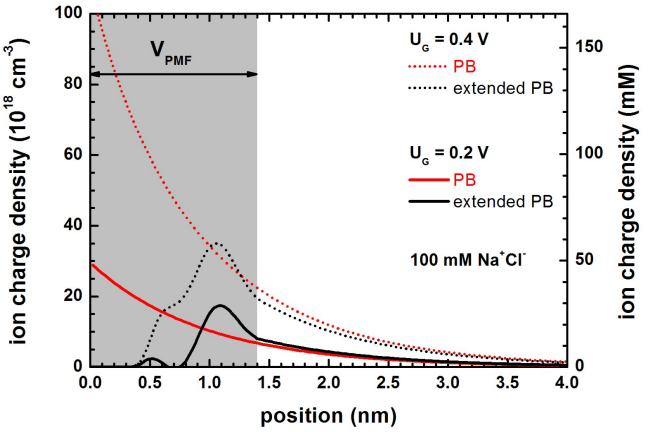


Figure 9.9.: Total ion density profile $\rho(x)$ for applied gate potentials of $U_G = 0.2\,V$ (solid lines) and $U_G = 0.4\,V$ (dotted lines) for the hydrophobic solid–liquid interface (extended Poisson–Boltzmann model, black lines) and for the standard Poisson–Boltzmann approach (red lines). The gray rectangle indicates the region where the potentials of mean force are nonzero. The 0.34 nm wide region of zero charge at the solid–liquid interface in the case of the hydrophobic extended PB model corresponds to an effective 'insulator thickness'.

employed while assuming a constant value of $\varepsilon_r = 78$ for the static dielectric constant of the electrolyte. We find significant differences for both the spatial distribution of the resulting ion density and the electrostatic potential distribution. Figure 9.9 shows the calculated total ion density profile $\rho(x)$ for applied gate potentials of $U_G = 0.2\,V$ (solid lines) and $U_G = 0.4\,V$ (dotted lines) for the hydrophobic solid–liquid interface (extended Poisson–Boltzmann model, black lines). The results of the standard Poisson–Boltzmann approach (red lines) are shown for comparison. The gray rectangle indicates the region where the potentials of mean force are nonzero. One can clearly see the effective 'insulator thickness', i.e. the 0.34 nm wide region of zero charge at the solid–liquid interface in the case of the hydrophobic extended Poisson–Boltzmann model. For higher gate voltages, the Poisson–Boltzmann approach leads to unrealistically high ion concentrations at the interface (not shown). We note that the integrated ion density in the electrolyte corresponds exactly to the sheet charge density in graphene due to the overall charge neutrality requirement.

Figure 9.10 shows the calculated electrostatic potential distribution for applied gate potentials $U_G = 0.2\,V$, $U_G = 0.4\,V$ and $U_G = 1.0\,V$ across the hydrophobic graphene–electrolyte interface (extended Poisson–Boltzmann, black solid lines). The results of the standard Poisson–Boltzmann approach (red dashed lines) are shown for comparison. In the latter case, the potential drop in the electrolyte is very small, so that the applied gate voltage nearly directly changes the Fermi level in the graphene layer with respect to the Dirac point. In the hydrophobic model, there is a significant potential drop in the electrolyte in the region close to the interface where the total ion density is very low

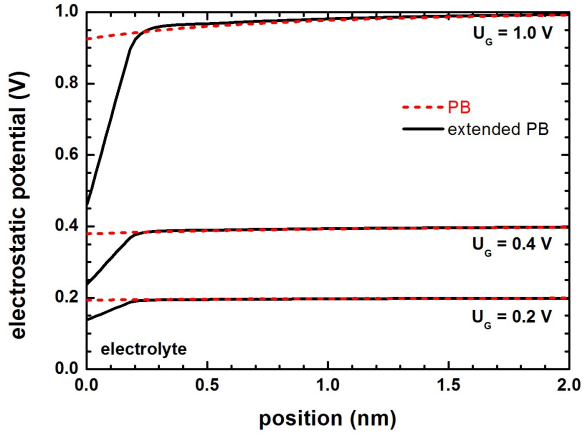


Figure 9.10.: Electrostatic potential distribution for applied gate potentials $U_G = 0.2\,V$, $U_G = 0.4\,V$ and $U_G = 1.0\,V$ across the hydrophobic graphene–electrolyte interface (extended Poisson–Boltzmann model, black solid lines). The results of the standard Poisson–Boltzmann approach (red dashed lines) are shown for comparison. In the hydrophobic model, there is a significant potential drop in the electrolyte in the region close to the interface where the total ion density is very low (see Fig. 9.9).

(see Fig. 9.9), resulting in a lower value of the electrostatic potential in the graphene layer. Consequently, for the same gate voltage the carrier density is expected to be lower in the hydrophobic model. Only results for positive U_G are shown. The results for negative U_G are symmetric with respect to the potential axis for the PB model, and also roughly symmetric for the extended PB model apart from very small deviations due to the different PMFs of the Na^+ and Cl^- ions.

We also find differences in terms of integrated charge densities (sheet densities) and capacitances among our two models. Figure 9.11 (left) shows the calculated carrier sheet density in the graphene layer as a function of applied gate electrode potential U_G for the hydrophobic solid–liquid interface using the extended Poisson–Boltzmann (ePB) approach (blue line). The results of the standard Poisson–Boltzmann (PB) approach (red line) are shown for comparison. The gray line shows the 'quantum limit' in bulk graphene where a shift in the applied voltage corresponds directly to a shift of the Fermi level with respect to the Dirac point, i.e. $E_F^{\text{graphene}} = eU_G$ The upper x axis corresponds to the experimental results of in-solution Hall effect measurements of Ref. [DHL+10] (blue dots), the lower x axis has been shifted so that the Dirac point is at $U_G = 0\,eV$. Close to the Dirac point, no experimental data is provided because it is difficult to estimate the number of carriers from Hall effect experiments in this case. The reason is that the Hall voltage goes to zero if the number of holes and electrons is similar. From Fig. 9.4 and Fig. 9.7 (arrow) the density around the Dirac point at room temperature is expected to be about $10^{11}\,cm^{-2}$. The experiment indicates a linear slope of the electron density vs. (positive) gate voltage and a supralinear dependence for the hole

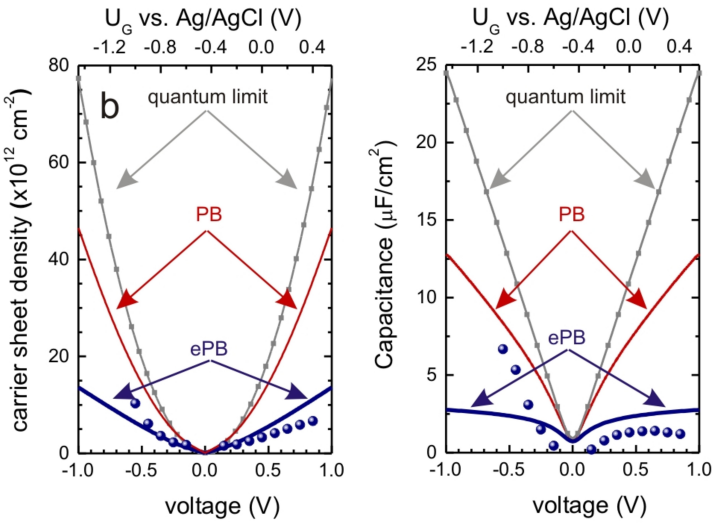


Figure 9.11.: Left: Calculated carrier sheet density in the graphene layer as a function of applied gate electrode potential U_{G} for the hydrophobic solid–liquid interface using the extended Poisson–Boltzmann (ePB) model (blue line) and the standard Poisson–Boltzmann (PB) approach (red line). The gray line shows the quantum limit in bulk graphene where a shift in the applied voltage corresponds directly to a shift of the Fermi level with respect to the Dirac point. The upper x axis corresponds to the experimental results (blue dots), the lower x axis has been shifted so that the Dirac point is at $U_{\mathrm{G}} = 0\,\mathrm{eV}$. Right: Calculated capacitance $C = \partial Q/\partial U_{\mathrm{G}}$ from the same data. In each graph, the left half corresponds to holes, and the right half to electrons. The blue dots are experimental data.

density vs. (negative) gate voltage. In the case of the PB model, the potential drop in the electrolyte is almost negligible (compare with Fig. 9.10), i.e. the electrostatic potential in graphene approximately equals U_{G}, resembling the 'quantum limit' situation. In the case of the extended PB model, an important potential drop occurs in the electrolyte, which strongly reduces the effective electrostatic potential in the graphene film and thus the carrier density. In any case, the graphene–electrolyte interface cannot be solely described by the quantum capacitance since the experimental carrier density is much smaller than predicted by the quantum limit. The right part of this figure shows the calculated capacitance $C = \partial Q/\partial U_{\mathrm{G}}$ from the same data. The latter is defined as the derivative of the total charge $Q = e\,(p - n)$ in the graphene sheet with respect to the applied gate potential U_{G}. In each graph, the left half corresponds to holes, and the right half to electrons. The qualitative agreement between theory and experiment is quite good for the electrons. The hydrophobic model leads to results reasonably close to measured electron densities in graphene. This also applies to the calculated capacitance as a function of voltage which shows a similar behavior than the experimentally obtained values.

However, our results for the hole densities were much lower compared to experiment. Thus we were not able to reproduce the measured asymmetry with respect to the electron and hole densities. Consequently, also the asymmetry in the capacitance plot could not be reproduced. We want to emphasize that we did not use any adjustable parameters for our calculations. Experimentally, also a clear asymmetry for electron and hole mobilities was found [DHL$^+$10], with the electron mobility being noticeably higher than the hole mobility for the same carrier concentration. The asymmetry in the charge densities in graphene could possibly be explained by ion (OH^-) adsorption at the graphene interface (similar as in diamond), which is only present at negative gate voltages [ACWL08]. This effect has not been taken into account in our simulations. In our model, the pH has been set to 7, thus the concentrations of OH^- and H_3O^+ ions are not relevant for the simulations compared to the Na^+ and Cl^- concentrations of 100 mM NaCl. The buffer ions (5 mM PBS buffer) have not been included in the simulations as they are nearly two orders of magnitude lower and potentials of mean force are not yet available for them. We also did not take into account the presence of charged impurities that can induce chemical doping, which would significantly increase the number of carriers by shifting the Fermi level away from the Dirac point in the absence of a gate potential. For simplicity, pyroelectric charges at the SiC–graphene interface have not been considered in our model. They arise due to the spontaneous polarization in hexagonal SiC–6H. A built-in electric field at the SiC–graphene interface could lead to electron doping of the graphene layer. Bilayer graphene has a different effective mass for electrons and holes. This would lead to asymmetry in the results for the carriers. It cannot be excluded with certainty that some parts on the experimental samples consisted of bilayer instead of single layer graphene.

Figure 9.12 shows the interfacial capacitance of the graphene–electrolyte system as a function of applied gate electrode potential U_G for the hydrophobic solid–liquid interface (extended Poisson–Boltzmann model, black solid line). The results of the standard Poisson–Boltzmann approach (red dashed line) are shown for comparison. The blue line shows the quantum limit of bulk graphene. These curves are the same as in Fig. 9.11 (right). The gray line corresponds to a simple plate capacitor model of width $d = 0.32\,\text{nm}$ and a static dielectric constant of $\varepsilon_r = 1$. One can see that at large voltages U_G the plate capacitor model describes nicely the hydrophobic double layer capacitance whereas for small voltages the quantum capacitance of graphene dominates. The reason for this behavior is simply the fact that the interfacial capacitance is a series capacitance of the quantum capacitance of graphene and the double layer capacitance of the electrolyte (eq. (9.9)). As shown in Fig. 9.6, potential fluctuations in graphene increase the quantum capacitance only at low voltages, thus increasing the interfacial capacitance only in this voltage regime. For the modeling of liquid-gated graphene in this section, we assumed no potential fluctuations, corresponding to an ideal graphene layer. The influence of these fluctuations can easily be estimated by comparing Fig. 9.12 with Fig. 9.6, i.e. only the low-voltage regime will be affected where the capacitance will increase slightly. Essentially, fluctuations will reduce the gate dependence of the interfacial capacitance, and may even lead to a rather constant value for the interfacial capacitance. Consequently, our results for ideal graphene are in fact the lower limit for

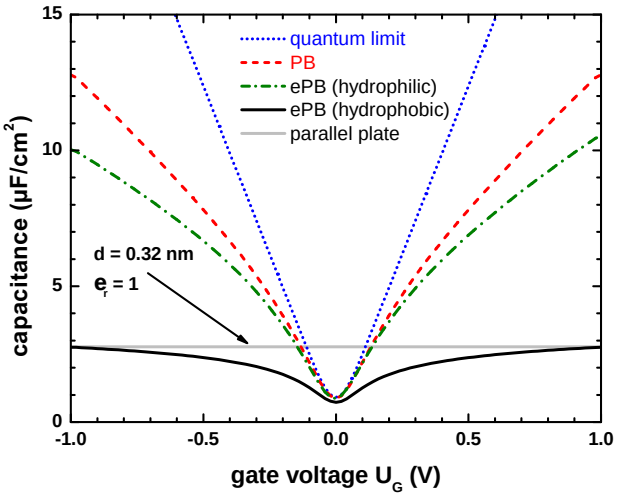


Figure 9.12.: Interfacial capacitance of the graphene–electrolyte system as a function of applied gate electrode potential U_{G} for the hydrophobic solid–liquid interface (extended Poisson–Boltzmann model, black solid line). The results of the standard Poisson–Boltzmann approach (red dashed line) and the quantum limit (blue dotted line) of bulk graphene are shown for comparison (same data as in Fig. 9.11). The gray line corresponds to a simple plate capacitor model of width $d = 0.32\,\mathrm{nm}$ and a static dielectric constant of $\varepsilon_{\mathrm{r}} = 1$. At large voltages U_{G} the plate capacitor model describes nicely the hydrophobic double layer capacitance whereas for small voltages the quantum capacitance of graphene dominates. For completeness we also include the results for the hydrophilic extended PB model (green dash-dotted line).

the interfacial capacitance. Our value for the capacitance of the double layer (parallel plate model) is around 3 $\mu\mathrm{F/cm^2}$ and thus much lower than the value of 70 $\mu\mathrm{F/cm^2}$ as estimated by Heller et al., who significantly overestimated the double layer capacitance. Also the standard Poisson–Boltzmann model overestimates the capacitance. For completeness, we also include our results for the capacitance of a (hypothetically) hydrophilic graphene–electrolyte interface (green dash-dotted line). Here, we used the same extended Poisson–Boltzmann model as for the hydrophobic case but instead used the hydrophilic parameters for the water density, i.e. static dielectric constant of the electrolyte, and for the PMFs of the ions, see Fig. 8.8 and Fig. 8.13, respectively. The hydrophilic results are somewhat closer the PB model but significantly different from the hydrophobic model. As our hydrophobic model leads to results that are closer to experimental results than the hydrophilic or the simple PB model, it seems to be very important to consider the hydrophobic nature of interfaces when analyzing, optimizing and modeling device behavior. In the next chapter we test our hydrophobic model on a different material system, namely diamond.

10. Modeling diamond based solution gated field-effect transistors

In this chapter we are investigating diamond–electrolyte interfaces. First, we discuss the band structure of diamond. Then we analyze the properties of the two-dimensional hole gas in surface conducting diamond as a function of substrate orientation. Finally, we model the hydrophobic interaction and charge accumulation at the diamond–electrolyte interface.

Hydrogen-terminated diamond is known to be surface conducting and hydrophobic when in contact with water. While the hydrophobic interaction of surfaces with water is a well-known phenomenon, there is not much known on its influence on biosensor devices. In this chapter we calculate the interfacial potential at the hydrogen-terminated diamond–aqueous electrolyte interface. We show that experimental results on the sheet charge density of diamond field-effect devices can be reproduced by our simulations, only if we include the hydrophobic nature of the surface into our model. Thus the performance of potentiometric biosensor devices strongly depends on the hydrophobicity of the surface. In Ref. [DLB$^+$11] we published our work on modeling the diamond–electrolyte interface where we particularly focused on the hydrophobicity of diamond and the charge accumulation in this electric double layer as a function of electrode gate potential. The hole charge distribution in the diamond had been described successfully with a single-band effective-mass model, although such a model is typically not appropriate to describe the hole energy levels in a semiconductor. However, for the purpose of that work where only the total charge as a function of gate potential was relevant, the single-band model worked reasonably well. In this chapter we extend the previous approach [DLB$^+$11], by now using a 6×6 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian to describe the charge accumulation at the diamond surface. To our knowledge this is the first attempt to apply a self-consistent $\mathbf{k} \cdot \mathbf{p}$ formalism to a semiconductor–electrolyte structure, where nonparabolicity effects and warping are included. Self-consistent solutions of the Schrödinger–Poisson equation for hydrogen-terminated diamond (surface conducting diamond) have previously been performed by Edmonds et al. [EPL10]. They considered a single-band model with effective masses derived from the Luttinger parameters of Willatzen et al. [WCC94] for diamond. In contrast to the single-band model with parabolic and isotropic masses, the $\mathbf{k} \cdot \mathbf{p}$ model allows us to compare different diamond substrate orientations, namely (100), (110) and (111) with respect to their sensitivity. However, an important ingredient for $\mathbf{k} \cdot \mathbf{p}$ calculations are the Luttinger parameters that describe the hole masses. It seems that there is not much known on the precise values of the valence band masses in diamond. This will be the topic of the next section.

10.1. Band structure of diamond

Intrinsic diamond is an insulator with an indirect band gap of 5.5 eV at room temperature. In this section we discuss the valence band structure of diamond and compare the energy dispersion along several directions in $\mathbf{k}$ space. For the $\mathbf{k} \cdot \mathbf{p}$ energy dispersions a number of different sets of Luttinger parameters that can be found in the literature will be tested against each other. Surprisingly, there is still a substantial lack of information about the details of the band structure of diamond. Willatzen et al. [WCC94] collected eight different sets of Luttinger parameters from different authors. Further comparisons have been made by Gheeraert et al. [GKTK99] and Reggiani et al. [RWZ83]. So we have at least 13 different sets of Luttinger parameters. Only two of them are similar (the ones of Saslow et al. and van Vetchen et al.), all others deviate more or less substantially. Previous work did not compare the resulting valence band structures for these parameters. We found that for actually four of these sets of Luttinger parameters the hole dispersion bends into the opposite direction (negative mass). Figure 10.1 shows our results. We only plot the energy dispersion along the directions [110] (solid lines) and [111] (dotted lines) (left part of the figure) where the sets of parameters lead to incorrect curvature of the hole bands. The dispersion along [100] is plotted in all cases (right part of the figure, solid lines). In diamond the split-off energy is very small ($\Delta_{\mathrm{so}} = 6\,\mathrm{meV}$) and indicated by the arrow. As we discard these sets of Luttinger parameters in the following, we do not list the actual Luttinger parameters and their references. Instead we refer to Table IV

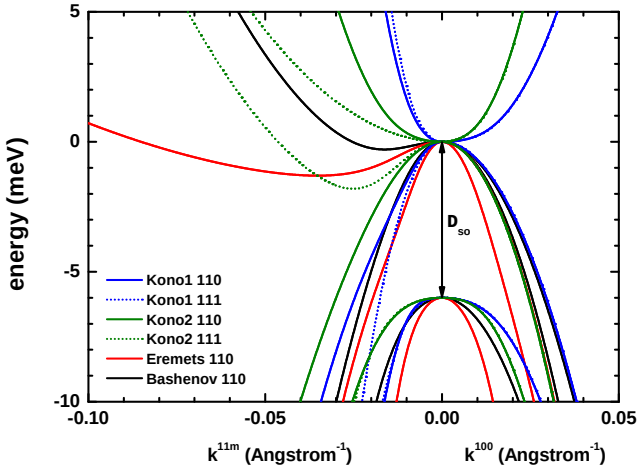


Figure 10.1.: Comparison of different sets of Luttinger parameters where the 6×6 $\mathbf{k} \cdot \mathbf{p}$ valence band energy dispersion leads to incorrect curvature of the hole bands. Negative x axis: energy dispersion along the [110] ($m = 0$, solid lines) and [111] ($m = 1$, dotted lines) directions; positive x axis: energy dispersion along the (100) direction (solid lines).

in Ref. [WCC94] and references therein. The parameters by Eremets et al. (black solid lines) and Bashenov et al. (red solid lines) lead to a dispersion along the [110] direction where the curvature is almost flat and has negative slope at $|k| > 0.035\,\text{Å}^{-1}$ (Eremets) or at $|k| > 0.016\,\text{Å}^{-1}$ (Bashenov). This fact was already discussed by Bashenov. Kono et al. have published two sets of parameters. For both the first set (blue lines) and the second set (green lines), the dispersion of the uppermost hole state has negative slope along all three directions [100], [110] and [111]. Furthermore, the second set (green dotted line) leads to negative slope for values of $|k| > 0.025\,\text{Å}^{-1}$ along the [111] direction. This makes us believe that not all authors actually calculated the band structure for their set of Luttinger parameters. Luttinger parameters are usually determined by cyclotron resonance experiments. This was the method employed by Rauch [Rau61] who derived an effective light hole mass of $m_{\text{lh}} = 0.70m_0$ that showed little anisotropy and a split-off hole mass of $m_{\text{so}} = 1.06m_0$ with no anisotropy. He also derived a value for the heavy hole mass of $m_{\text{hh}} = 2.18m_0$ [Rau62]. We used these values in our previous work [DLB$^+$11] where we employed a single-band model. We note that the A, B and C parameters of the early work of Rauch are significantly different to all other sets of Luttinger parameters, and there seems to be even doubt on the correct interpretation of his experiment [GKTK99].

Figure 10.2 and Fig. 10.3 show the dispersion along [110] and [111], respectively, as well as along [100] for other sets of Luttinger parameters. We did not include the sets published by van Haeringen et al., Reggiani et al., Saslow et al., van Vetchen et al. and Hall. The purpose of the figures is to demonstrate that there is significant variance among those parameters, and that the energy dispersion of the heavy hole for the Rauch parameters corresponds to a much larger, i.e. heavier hole mass than for other sets of Luttinger parameters.

As there is currently no consensus about any experimentally derived set of Luttinger parameters, we also investigated the band structure obtained from the $sp^3d^5s^*$ tight-binding (TB) parameterization of Jancu et al. [JSBB98]. They published empirical TB parameters for diamond. However, this set does not lead to a spin-orbit splitting at $\mathbf{k} = 0$ and thus all three hole bands are degenerate at the Γ point. The reason for this is that they set their $\Delta/3$ parameter to zero. We adjusted this parameter to $\Delta/3 = 0.0024\,\text{eV}$ so that we now obtain a split-off energy of $\Delta_{\text{so}} = 0.006\,\text{meV}$. This change is negligible with respect to the overall band structure so that we can still use the original Jancu parameterization for all other parameters. Only the valence band structure at the Γ point is affected by our choice of $\Delta/3$. We compared the $\mathbf{k} \cdot \mathbf{p}$ dispersions for all sets of Luttinger parameters to our calculated tight-binding band structure (not shown). The Luttinger parameters of Willatzen et al. were the ones that were closest to the TB valence band structure along all three high symmetry directions as shown in Fig. 10.4. This is the reason why we use Willatzen's parameters in the following although we do not claim that this is the 'best' choice. Their parameters have been derived within the framework of density-functional theory (local density approximation, self-consistent scalar-relativistic linear muffin-tin-orbital method). Having established our set of Luttinger parameters we are now prepared to analyze diamond heterostructures in general or diamond surfaces as a function of substrate orientation in the next section.

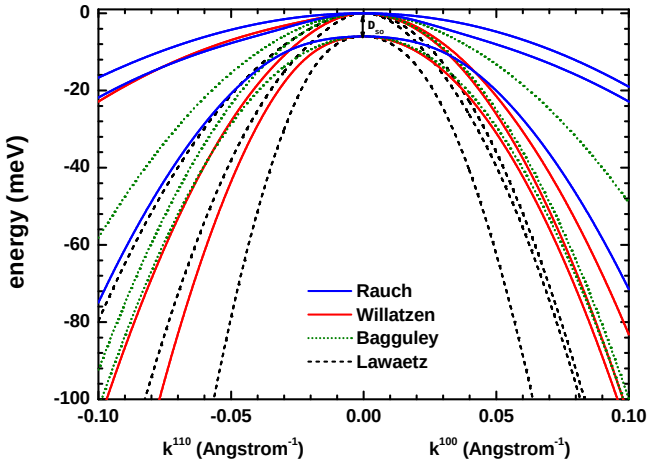


Figure 10.2.: 6×6 $\mathbf{k}\cdot\mathbf{p}$ valence band energy dispersion along the [110] and [100] directions for different sets of Luttinger parameters.

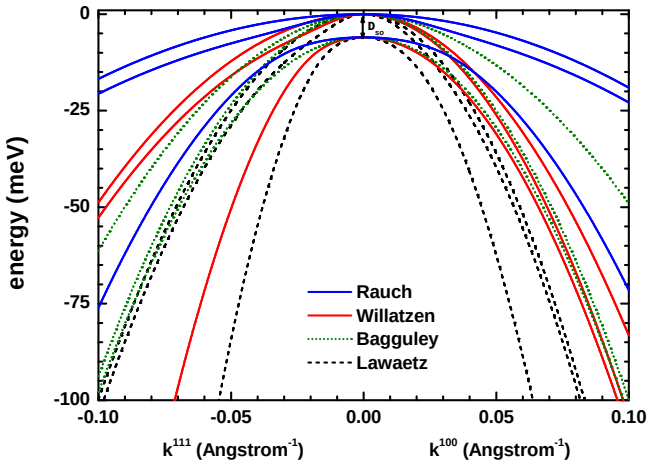


Figure 10.3.: Same as Fig. 10.2 but along the [111] and [100] directions.

10.2. Results: Influence of substrate orientation on the density of a two-dimensional hole gas in diamond

In this section we investigate the density of a two-dimensional hole gas (2DHG) in diamond for different substrate orientations at room temperature. Our model system consists of diamond where the 2DHG is induced by a constant surface charge density. In

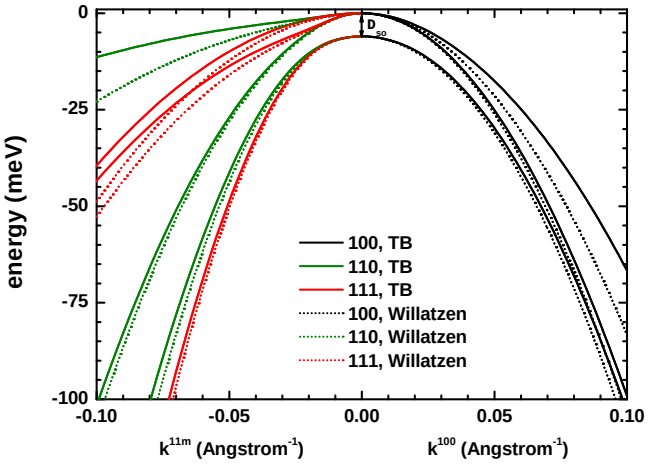


Figure 10.4.: Energy dispersion along the high symmetry directions [100] (black), [110] (green), [111] (red) in $\mathbf{k}$ space calculated by the tight-binding (TB) method (solid lines). For comparison the results obtained by diagonalizing the bulk 6×6 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian for each $\mathbf{k}$ vector is shown for the Luttinger parameters of Willatzen et al. (dotted lines).

real samples the 2DHG is induced by the hydrogen termination of the diamond surface. This results in a negative electron affinity of about $\chi = -1\,\mathrm{eV}$ which causes p-type surface conductivity (surface conducting diamond). The hydrogen at the diamond surface has another effect, namely that the surface gets hydrophobic. This will be the topic of the next section. In this section we solve the 6×6 $\mathbf{k} \cdot \mathbf{p}$ Schrödinger–Poisson equation self-consistently for (100), (110) and (111) substrate orientations. As boundary conditions we used a negative interface charge density of $\sigma = -5 \cdot 10^{13}\,\mathrm{cm}^{-2}$ at the diamond surface in order to induce a 2DHG, and a flat band boundary condition in the bulk diamond, i.e. zero electric field far away from the interface. The doping concentration was assumed to be n-type in the whole diamond ($0.5 \cdot 10^{18}\,\mathrm{cm}^{-3}$, nitrogen with ionization energy $E_\mathrm{D}^\mathrm{ion} = 1.7\,\mathrm{eV}$).

Figure 10.5 shows our results. It can be clearly seen that the (111) substrate orientation (red dashed lines) has a higher density than the (100) orientation (black solid lines). Even a higher density is obtained for the (110) orientation (blue dash-dotted lines). The corresponding 2DHG sheet densities are $\sigma_{110} = 5.8 \cdot 10^{12}\,\mathrm{cm}^{-2}$, $\sigma_{111} = 4.5 \cdot 10^{12}\,\mathrm{cm}^{-2}$ and $\sigma_{100} = 3.0 \cdot 10^{12}\,\mathrm{cm}^{-2}$, respectively. The different results for each orientation are due to the anisotropy of the valence band structure. It can also be seen from the sheet densities that the interface charge σ is not completely canceled (screened) by the 2DHG sheet charge density. Thus there is additional band bending further away from the 2DHG region and not a flat band. This also reveals the difference between the capacitance of an ideal parallel plate capacitor and the 'quantum capacitance' [Lur88] of a two-dimensional

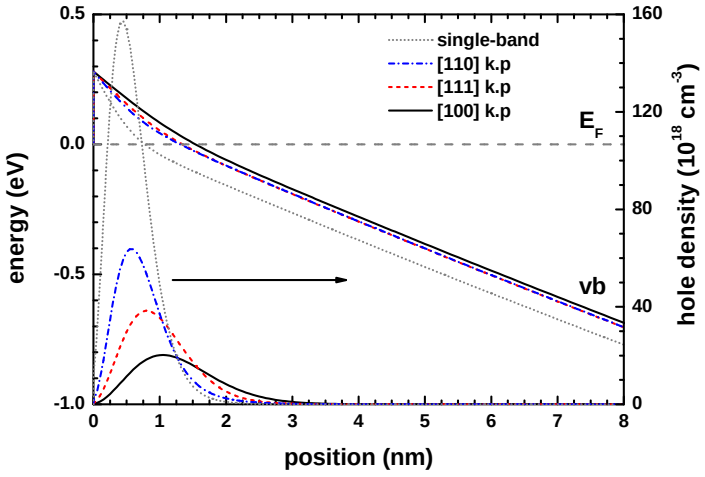


Figure 10.5.: Hole densities and valence band edges (vb) of a two-dimensional hole gas in diamond for various substrate orientations ((110) – blue dash-dotted lines, (111) – red dashed lines, (100) – black solid lines). The single-band results are independent of substrate orientation (gray dotted lines) because isotropic masses are used. The Fermi level E_F is indicated by the gray dashed line.

electron or hole gas. We used the Luttinger parameters of Willatzen et al., motivated by the discussion of the previous section. The parameters by Reggiani et al., Saslow et al., van Haeringen et al. and van Vetchen et al. show a higher anisotropy (warping) in terms of [111] vs. [100] directions leading to a slightly larger difference with respect to the respective densities in this example (not shown). For comparison we also show the single-band results (gray dotted lines) obtained with the parabolic and isotropic effective masses by Rauch ($m_\mathrm{hh} = 2.18m_0$, $m_\mathrm{lh} = 0.70m_0$, $m_\mathrm{so} = 1.06m_0$) where for each of the three valence band edges the single-band Schrödinger equation was solved. In this case the results are independent of substrate orientation. As these masses are much heavier than the masses by Willatzen (see also Fig. 10.2 and Fig. 10.3), the density is larger than for the $\mathbf{k}\cdot\mathbf{p}$ formalism. In fact, the single-band density is almost entirely due to the occupation of the highest heavy hole ground state. Also for $\mathbf{k}\cdot\mathbf{p}$ along [110] only the ground state contributes to the density. This is consistent as the mass along [110] is much heavier than along the other directions (see Fig. 10.4). For [111] the ground state and the first excited state contribute to the density because they have very similar energies, whereas for [100] the ground state and the first two excited states contribute to the density. The latter two excited states also have very similar energies. This can be seen in Fig. 10.6 which shows the three uppermost eigenstates (square of the probability amplitude shifted by its eigenenergy) for each substrate orientation at $\mathbf{k}_\parallel = 0$. In fact because of spin the six highest eigenstates are shown but the two spin states have the

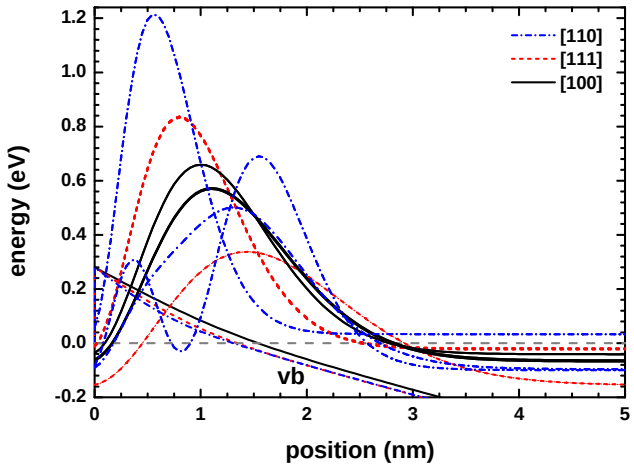


Figure 10.6.: Three uppermost eigenstates (square of the probability amplitude shifted by its eigenenergy) for each orientation at $\mathbf{k}_{\|} = 0$, valence band edge energies (vb) and Fermi level.

same energy and the same probability density at $\mathbf{k}_{\|} = 0$. Only for [110] (blue dash-dotted lines) the energy of the ground state lies above the Fermi energy (gray dashed line), compare also with Fig. 10.10. The probability densities of the uppermost three states have only one maximum for [100] (black solid lines) and [111] (red dashed lines) directions, i.e. they are derived from the ground states of the heavy, light and split-off hole but these are in fact mixed states. Only for [100] at $\mathbf{k}_{\|} = 0$, the second state is a pure heavy hole, and for [111] the ground state is 50% heavy and 50% light with no contribution from split-off hole at $\mathbf{k}_{\|} = 0$. All other states shown in this figure are mixed states. For [110] the second excited state has one node. For all orientations our calculations indicate that states with one node are not occupied (not shown). For confinement along [110] direction, the maximum of the probability density is much closer to the surface. This will have influence on the capacitance of diamond biosensors as the 2DHG density is then more sensitive to potential changes at the surface. In the figure it looks as if only the first two states are plotted for [100] and [111] but in fact three states are plotted for each. The reason is that for [100] the two excited states have almost the same energy (separated by 4 meV) and the same shape, whereas for [111] the ground state and the first excited state have almost the same energy (also separated by 4 meV) and the same shape (see also Fig. 10.10 where the energies are shown in more detail). As the triangular-like confinement potential is very strong, the 2DHG is located within the first few nanometers. This justifies to use a small quantum region of 10 nm (nonuniform grid, 151 quantum grid points) with Dirichlet boundary conditions. The $\mathbf{k}_{\|} = (k_x, k_y)$ space has been discretized on $41 \times 41 = 1681$ $\mathbf{k}$ points with a maximum value of $k_{\max} = 0.18\,\text{Å}^{-1}$ along the k_x and k_y directions.

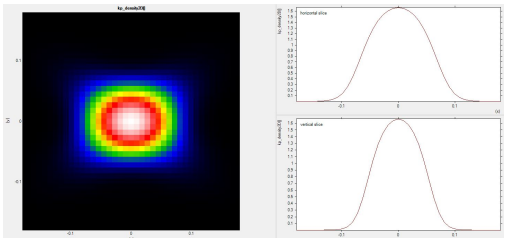


Figure 10.7.: $\mathbf{k}_{\parallel}$-resolved hole density distribution $p(k_x, k_y)$ for (110) oriented diamond. Here, k_x is related to [100] and k_y to $[01\overline{1}]$ direction. The confinement direction is parallel to [011]. The right part shows the horizontal and vertical slice through the center.

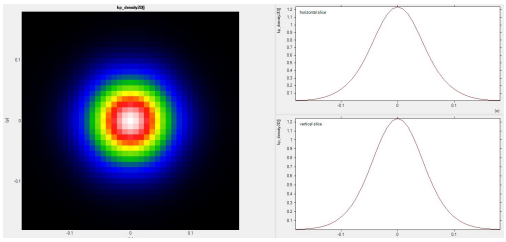


Figure 10.8.: $\mathbf{k}_{\parallel}$-resolved hole density distribution $p(k_x, k_y)$ for (111) oriented diamond. Here, k_x is related to $[11\overline{2}]$ and k_y to $[\overline{1}10]$ direction. The confinement direction is parallel to [111].

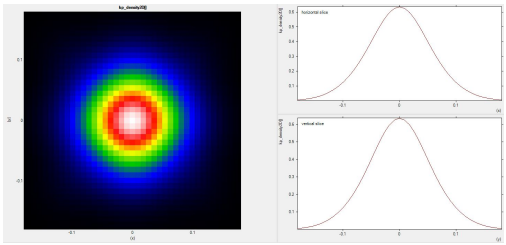


Figure 10.9.: $\mathbf{k}_{\parallel}$-resolved hole density distribution $p(k_x, k_y)$ for (100) oriented diamond Here, k_x is related to [100] and k_y to [010] direction. The confinement direction is parallel to [001].

Figure 10.7, Fig. 10.8 and Fig. 10.9 show the $\mathbf{k}_\parallel$-resolved hole density distribution $p(k_x, k_y)$ for (110), (111) and (100) oriented diamond, respectively. In these plots one can see the amount that each $\mathbf{k}_\parallel$ point contributes to the density. The $\mathbf{k}_\parallel = (k_x, k_y)$ space refers to the rotated coordinate system and thus the k_x and k_y direction are different for each orientation (see figure captions). In all cases the growth direction is meant to be along the z direction. The $\mathbf{k}_\parallel$-resolved hole density distribution for (100) oriented diamond has to be symmetric with respect to k_x and k_y (Fig. 10.9). The $\mathbf{k}_\parallel$-resolved hole density distribution for (111) looks symmetric with respect to k_x and k_y (Fig. 10.8). A detailed analysis of the energy dispersion (see Fig. 10.10) reveals that the dispersion along k_x is in fact very similar to the one along k_y (red dotted lines), thus justifying to expect an almost symmetric $\mathbf{k}_\parallel$-resolved density. This is not the case for (110) oriented diamond (Fig. 10.7). In this case the $\mathbf{k}_\parallel$-resolved hole density distribution is elongated along the k_x direction indicating that it is more favorable to occupy states along k_x rather than along k_y. To understand this preference it is necessary to analyze the energy dispersion of the occupied states. Figure 10.10 reveals that the energy dispersion of the ground state h_1 (blue dash-dotted line) has less curvature along the k_x direction rather than the k_y direction. Consequently, this favors the occupation of the states along the k_x direction for increasing energy.

Finally, we show the calculated energy dispersions of the highest hole eigenstates for each substrate orientation in Fig. 10.10. The energy scale is the same as for Fig. 10.6 which shows the eigenstates at $\mathbf{k}_\parallel = (k_x, k_y) = 0$. Only for (110) oriented diamond the ground state h_1 lies above the Fermi level E_F. The hole energy dispersion for (100) oriented diamond is symmetric with respect to k_x and k_y whereas for (111) diamond a slight asymmetry is present. (110) diamond shows a very pronounced asymmetry as already discussed. Also the energies at $\mathbf{k}_\parallel = 0$ and their contributions to the density have already been discussed when analyzing the square of the probability amplitudes (Fig. 10.6). At $\mathbf{k}_\parallel = 0$ each state is twofold spin degenerate. Thus we labeled the uppermost 'three' states with h_1, h_2 and h_3 although there are actually six states. The spin-splitting at nonzero $\mathbf{k}_\parallel$ is due to structural inversion asymmetry (SIA). The triangular confinement potential induces SIA in contrast to bulk diamond which has a center of inversion, i.e. no bulk inversion asymmetry (BIA) and no SIA. Only the spin-splitting of the states h_1 and h_2 for (100) diamond is noticeable in this figure, in all other cases it can hardly be recognized.

Our calculations reveal a lot of information in terms of energies, energy dispersions, spatial extension and character of wave functions for each $\mathbf{k}_\parallel$ vector, as well as information on subband densities or even $\mathbf{k}_\parallel$-resolved densities for each substrate orientation. This information is useful in understanding and optimizing device designs, although we emphasize that reliable band structure parameters are needed as input. Our analysis is extremely useful when transitions between these hole subbands are involved, e.g. in optical absorption experiments, as different orientations show a significantly different energy spectrum, e.g. due to selection rules. However, for experiments where mainly the density is involved, e.g. in capacitance–voltage measurements, only the total density is relevant. Thus the contribution of the individual subband densities to the total density is somehow irrelevant, i.e. the precise spectrum of the energy levels is not as important

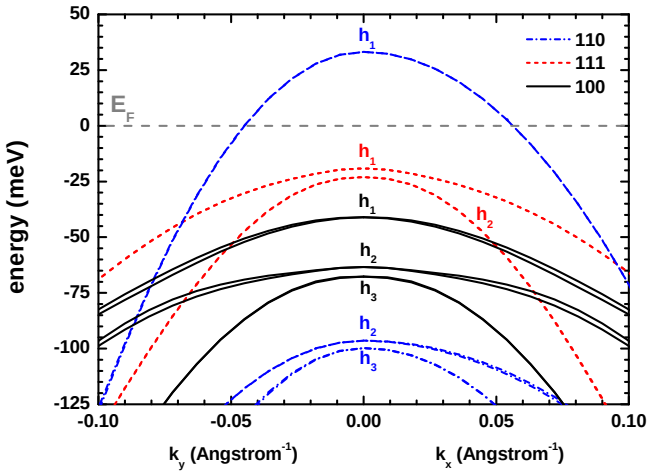


Figure 10.10.: Dispersion of the highest hole energy levels for (100), (110) and (111) oriented diamond. Only for (110) oriented diamond (blue dash-dotted lines) the ground state lies above the Fermi level. The k_x and k_y directions are with respect to the rotated coordinate system and thus are different for each orientation. (100) oriented diamond (black solid lines) is symmetric with respect to k_x and k_y whereas (111) diamond (red dashed lines) shows a slight asymmetry. (110) diamond shows a very pronounced asymmetry.

as for optical absorption experiments. Depending on the orientation either one or several subbands are occupied but this is not vital as here merely the total density is relevant.

Neglecting any additional surface effects like surface relaxation or surface reconstruction, our calculations demonstrate that a (110) sample has a higher 2DHG sheet density than a (111) sample, and that the latter has a higher sheet density than a (100) sample with respect to the same boundary condition (negative interface charge). This is equivalent to saying that the change in 2DHG sheet density of a (110) sample is larger with respect to potential changes at the surface than for a (111) or for a (100) sample, i.e. a (110) sample is more sensitive than (111) and (100) samples, and consequently better suited for sensor devices. The change of 2DHG sheet density due to potential changes will be discussed in more detail in the next section where the surface potential is varied by changing the potential in an electrolyte.

10.3. Results: Hydrophobic interaction and charge accumulation at the diamond–electrolyte interface

Biosensor and bioelectronic devices are still a matter of intensive research and their commercial success has remained a challenge. Interfaces with water play a major role in

these devices. Most biosensors are based on silicon where the silicon layer has a native thin oxide layer between the silicon and the electrolyte. This is a similar arrangement as in MOSFETs where the silicon layer is separated from the gate by an oxide which acts as an insulator. Analogous to a plate capacitor, the potential-dependent charge in the conductive channel of a field-effect transistor constitutes the capacitance of the interface. Consequently, the influence of the oxide on the capacitance is well studied in silicon. To increase silicon CMOS device performance the thickness of this oxide layer has been reduced over the last decades, but also SiO_2 has been replaced with high-k dielectric materials like HfO_2 with a large dielectric constant of $\varepsilon_r = 26.1$ (see Fig. A.3). The dominance of silicon in biosensor applications is being challenged by new materials like nitrides, graphene (see Chapter 9), carbon nanotubes or diamond due to their better biocompatibility, chemical and electrochemical stability in electrolytes and functionalization possibilities. The carbon based materials do not have a native oxide and thus the two-dimensional carrier gas is very close to the electrolyte which acts as a gate. One can thus assume that the sensitivity is higher than in silicon based devices. In most biosensors the sensing signals are generated by potential changes across the interface that modulate the concentration of the charge carriers, and thus the conductivity. We will show in the following that the strong hydrophobic nature of the diamond interface will reduce the sensitivity in comparison to hydrophilic interfaces. This is due to the fact that the hydrophobicity of the surface leads to a depletion of water in the vicinity of the surface. Therefore this hydrophobic 'gap' acts as a small insulating layer of very low dielectric constant which influences the total capacitance, i.e. in this very narrow region the dielectric constant drops from around $\varepsilon_r = 78$ to approximately $\varepsilon_r = 1$. An additional effect that influences the capacitance is the position of the ion charges in the electrolyte relative to the two-dimensional carrier gas. For hydrophobic interfaces the ions are further away than for hydrophilic interfaces. These effects have been barely discussed in the literature when explaining the operation of biosensor devices, but in fact the hydrophobicity of a surface has a major effect on its interaction with water. In the more commonly employed Si–SiO_2 based biosensors such effects are negligible. First, because their surface is hydrophilic, and secondly because the oxide acts as a dielectric spacer between the charges in the solid and in the liquid. Thus any effects of additional hydrophobic separation would be hardly noticeable. We note that organic semiconductors also exhibit strong hydrophobicity. Due to the electrochemical stability of the diamond surface, it is possible to bring it into an electrolyte solution and to apply a potential between the electrically contacted diamond surface and a reference electrode in the electrolyte. The potential across the diamond–electrolyte interface determines the position of the band edges relative to the Fermi level at the diamond surface. If the valence band edge is pushed close to or above the Fermi level, holes are accumulated and form a two-dimensional hole gas at the diamond surface. Its charge carrier density is determined by the position of the valence band edge relative to the Fermi level. Thus the hole density at the diamond surface increases with applied voltage where the capacitance of the interface determines the respective part of the voltage drop in the diamond and in the electrolyte. In the following we will model this potential drop by calculating the electrostatic potential distribution across the interface and its corresponding charge

accumulation by solving the coupled system of the 6×6 $\mathbf{k} \cdot \mathbf{p}$ Schrödinger and the Poisson equation self-consistently. The Schrödinger equation is solved only in the diamond but the Poisson equation is solved in both the diamond region and in the electrolyte region. The charge density consists of the density of all ions in the electrolyte and the hole density in the diamond determined by the wave functions and the energy levels as obtained by the Schrödinger equation. We use a Dirichlet boundary condition for the electrostatic potential in the bulk electrolyte ($\phi(\infty) = U_\mathrm{G}$) and a Neumann boundary condition ($\frac{\partial \phi}{\partial x} = 0$) in the bulk diamond at the left side. U_G is the gate potential applied between the reference electrode and the diamond contact. We will compare three different models, the standard Poisson–Boltzmann (PB) approach and the extended Poisson–Boltzmann approach assuming a hydrophobic or a hydrophilic interface. The latter two models take into account a spatially varying dielectric constant $\varepsilon_\mathrm{r}(x)$ and the potential of mean force (PMF) for the electrolyte ions (see Section 8.4), whereas the PB model assumes a homogeneous dielectric constant of water ($\varepsilon_\mathrm{r} = 78$) in the whole electrolyte region up to the interface, and no PMFs. The electrolyte consists of 50 mM NaCl. As we do not have available the PMFs for the 10 mM potassium based phosphate buffer ions (PBS, see Subsection 8.3.5), we simply ignore the buffer ions in the following.

In the previous section we found differences among the three diamond substrate orientations. Experiments on (110) diamond are difficult because (110) oriented diamond samples are hardly available. While studying the electronic transport at the hydrogen-terminated diamond–electrolyte interface, Lippert found experimentally that (111) diamond is more sensitive than (100) diamond [Lip10]. This was qualitatively confirmed by our diamond–electrolyte simulations (not shown) but the increase in sensitivity was not as large as expected. Therefore we cannot fully explain the difference in experimental results between (111) and (100) diamond. We thus believe that the increased sensitivity of the (111) oriented diamond sensor cannot be explained only by the different hole band structure along this direction. It seems reasonable that also the different atomic arrangement at the surface has a significant effect, e.g. surface relaxation or surface reconstruction, or that the (111) surface is slightly less hydrophobic than the (100) surface. In the following we present our results on the calculations of diamond–electrolyte interfaces where the diamond surface is oriented in the (100) plane.

In order to quantitatively understand the modulation of the 2DHG density by varying the electrolyte potential, we have to analyze the charge distribution at the solid–liquid interface in more detail. Figure 10.11 shows the calculated valence band edge energy (black lines) of the hydrogen-terminated diamond–electrolyte interface for an applied gate voltage $U_\mathrm{G} = -0.2\,\mathrm{V}$. By adjusting the gate potential applied between the reference electrode in the electrolyte and a contact on the diamond, one can shift the valence band edge with respect to the Fermi level. This modifies the confinement potential of the resulting triangular well, and thus also the positive charge density (blue lines) of the two-dimensional hole gas in the diamond, leading to a band bending at the surface. The hole density is mirrored by the corresponding total ion charge density (net negative charge) in the electrolyte indicated in red. The results of the Poisson–Boltzmann (PB) calculation are shown in dotted lines whereas the results obtained with our new extended Poisson–Boltzmann model that takes into account the hydrophobic interface are shown

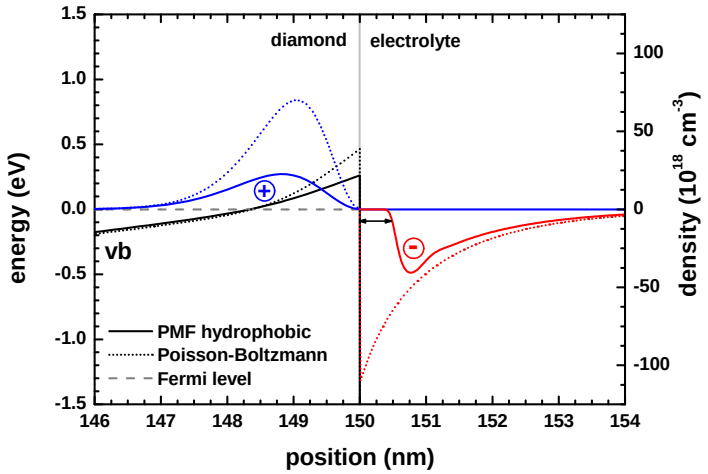


Figure 10.11.: Calculated valence band edge energy (black lines) of the hydrogen-terminated diamond–electrolyte interface for an applied gate voltage $U_G = -0.2\,V$. By adjusting the gate potential applied between the reference electrode in the electrolyte and a contact on the diamond, one can shift the valence band edge with respect to the Fermi level ($E_F = 0\,eV$ in diamond). This modifies the confinement potential of the resulting triangular well, and thus also the positive charge density (blue lines) of the 2DHG in the diamond, leading to a band bending at the surface. The hole density is mirrored by the corresponding total ion charge density (net negative charge) in the electrolyte (red lines). The results of the Poisson–Boltzmann calculation are shown in dotted lines whereas the results obtained with our new extended Poisson–Boltzmann model that takes into account the hydrophobic interface are shown in solid lines. The arrow indicates the region of low water density (hydrophobic region).

in solid lines. The arrow indicates the region of low water density (hydrophobic region) which has a width of approximately 0.3 nm. The same situation is shown in Fig. 10.12. In the left part the three highest eigenstates (square of the probability amplitude shifted by its eigenenergy) at $\mathbf{k}_{\parallel} = 0$ are shown for both the hydrophobic (red solid lines) and the standard PB approach (blue dotted lines). In fact, at $\mathbf{k}_{\parallel} = 0$ the states are twofold degenerate due to spin, so essentially six states are shown. In the right part the distribution of the Cl^- and the Na^+ ions is shown for both models. In the bulk electrolyte both ions reach their equilibrium concentration of 50 mM. The potential of mean force (PMF) for the Na^+ ions causes the local maximum in the Na^+ ion density profile. For the PB model both the 2DHG and the Cl^- ions are closer to the interface. In this case also higher densities are present on both sides (compare with Fig. 10.11). In the diamond part, this is achieved by moving the energy levels closer to the Fermi level.

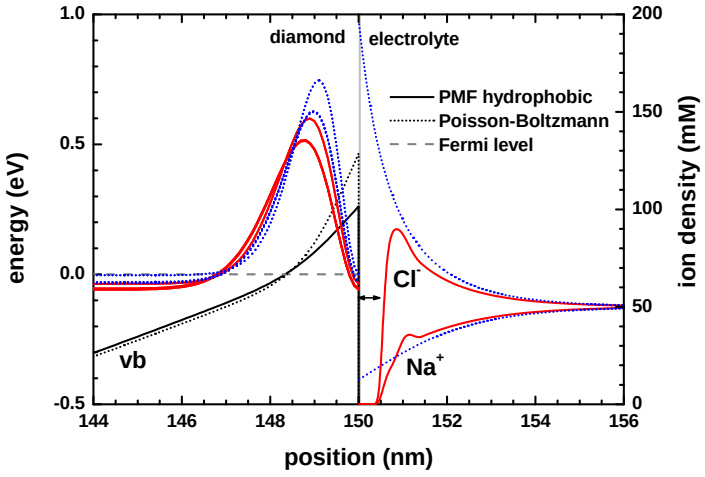


Figure 10.12.: Calculated valence band edge energy (black lines) for an applied gate voltage $U_G = -0.2\,V$, i.e. same situation as in Fig. 10.11. The arrow indicates the region of low water density (hydrophobic region). In the left part the three highest eigenstates (square of the probability amplitude shifted by its eigenenergy) at $\mathbf{k}_{\|} = 0$ are shown for both the hydrophobic (red solid lines) and the standard Poisson–Boltzmann (PB) approach (blue dotted lines). In the right part the distribution of the Na^+ and Cl^- ions is shown for both models (solid lines: extended PB model with hydrophobic interface, dotted lines: PB calculation).

As already shown in Fig. 10.6 of the previous section, the second and the third eigenstate have almost the same energy (also separated by 4 meV for both the hydrophobic and the PB model) and the same shape for (100) oriented diamond. This is however difficult to see in this figure. The 2DHG sheet charge density for the hydrophobic model is $\sigma = 3.8 \cdot 10^{12}\,cm^{-2}$ and for the PB model $\sigma = 10 \cdot 10^{12}\,cm^{-2}$.

Figure 10.13 shows the calculated electrostatic potential and valence band edge energy (black lines) for an applied gate voltage $U_G = -0.2\,V$, i.e. same situation as in previous figures. Again, the results of the Poisson–Boltzmann (PB) calculation are shown in dotted lines, whereas the results obtained with our new extended Poisson–Boltzmann model that takes into account the hydrophobic interface are shown in solid lines. The applied gate voltage U_G in the electrolyte cannot be directly related to the electrostatic potential in the diamond because part of the applied voltage drops in the electrolyte region close to the interface. The electrostatic potential distribution reveals the potential drop across the diamond–electrolyte interface. We found that there is a striking difference between the two models. The arrow indicates the large potential drop in the electrolyte region for the hydrophobic model (red solid lines). The potential drop in the electrolyte for the PB model (blue dotted lines) is much smaller because here the ions are allowed to approach

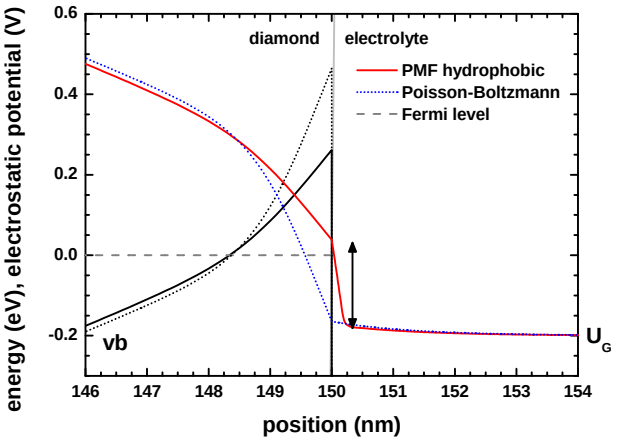


Figure 10.13.: Calculated electrostatic potential and valence band edge (black lines) of the hydrogen-terminated diamond–electrolyte interface for an applied gate voltage $U_\mathrm{G} = -0.2\,\mathrm{V}$, i.e. same situation as in previous figures. The arrow indicates the large potential drop in the electrolyte region for the hydrophobic model (red solid lines). The potential drop in the electrolyte for the standard Poisson–Boltzmann model (blue dotted lines) is much smaller. Consequently, the surface potential at the interface has almost the same value as the applied potential (solid lines: extended PB model with hydrophobic interface, dotted lines: PB calculation).

the interface infinitely close and additionally, the dielectric constant of water is very high even close to the surface. Both effects minimize the potential drop in the electrolyte. Consequently, the surface potential at the interface has almost the same value as the applied potential. In the latter case, most of the potential drop occurs in the diamond, whereas for the extended Poisson–Boltzmann model most of the potential drop happens in the hydrophobic 'gap' region where the ion concentration and the water density is very low. Recall that for the PB model the dielectric constant has a value of $\varepsilon_\mathrm{r} = 78$ up to the interface, whereas the extended PB model assumes a dielectric constant of $\varepsilon_\mathrm{r} = 1$ in the hydrophobic 'gap' where no ions are present. Consequently, this potential drop is very similar to the drop in the insulator region of a metal–insulator–semiconductor structure (see Appendix A). As a result, the valence band edge for the hydrophobic model is closer to the Fermi level resulting in a lower 2DHG density.

The experimental carrier concentrations [Lip10] for different gate potentials from in-liquid Hall effect measurements and the simulated results obtained with our $\mathbf{k}\cdot\mathbf{p}$ calculations are compared in Fig. 10.14. The experimental data sets (symbols) for several (100) diamond samples are normalized with respect to their threshold voltage. Above the threshold voltage all of them show a nearly linear increase of the carrier concentration with the gate potential U_G. The slope of the experimental data points is almost the same

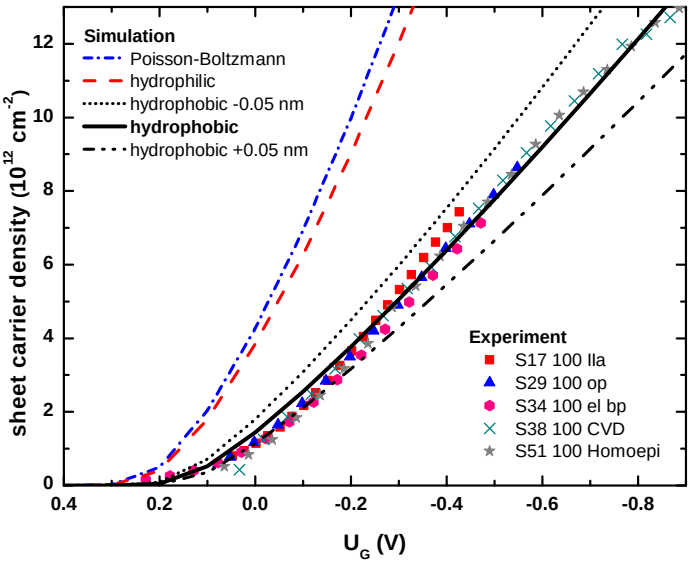


Figure 10.14.: Comparison of the experimentally obtained carrier concentrations for different gate potentials with our $\mathbf{k} \cdot \mathbf{p}$ calculations. The experimental data sets (symbols) for several (100) diamond samples are normalized with respect to their threshold voltage. The extended Poisson–Boltzmann model for the hydrophobic interface (black solid line), the hydrophilic interface (red dashed line) and the conventional Poisson–Boltzmann model (blue dash-dotted line) are compared. Results of the extended PB model for the hydrophobic case when a shift of the water density of ± 0.05 nm is considered are additionally shown (dash-dot-dot and dot).

for all samples. From the slope one can determine the interfacial capacitance to be about 2 μF/cm^2, see Fig. 10.15. This value is in agreement with results obtained from cyclic voltammetry and impedance spectroscopy [GNHS08]. The extended Poisson–Boltzmann model for the hydrophobic interface (black solid line), the hydrophilic interface (red dashed line) and the conventional Poisson–Boltzmann model (blue dash-dotted line) are compared. The agreement between the hydrophobic extended PB model and the slope of the experimental data is remarkable. We note that the hydrophobic model employed in our work is based on molecular dynamics simulations, see Section 8.4. In order to test the influence of the size of the water-depleted region, we have shifted the water density profile by a tiny amount of ± 0.05 nm with respect to the surface. This is a typical variation for different hydrophobic surfaces. These calculations (dash-dot-dot and dot) show that our assumed water density profile works surprisingly well. We also found that shifting the PMFs by this amount has only negligible influence (not shown).

Figure 10.15 shows the calculated capacitance–voltage characteristics of the extended Poisson–Boltzmann model for the hydrophobic interface (black solid line). Results of the extended Poisson–Boltzmann model for the hydrophobic case when a shift of the

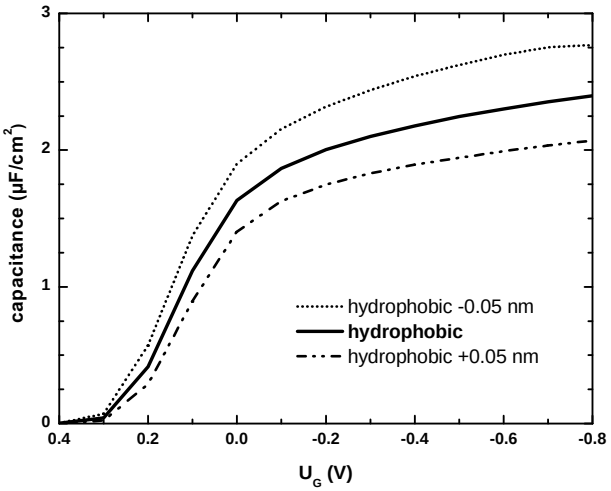


Figure 10.15.: Calculated capacitance–voltage characteristics of the extended Poisson–Boltzmann model for the hydrophobic interface (black solid line). Results of the extended Poisson–Boltzmann model for the hydrophobic case when a shift of the water density of $\pm 0.05\,\mathrm{nm}$ is considered are additionally shown (dash-dot-dot and dot).

water density of $\pm 0.05\,\mathrm{nm}$ is considered are additionally shown (dash-dot-dot and dot). The capacitance values of around 2 $\mu\mathrm{F/cm^2}$ are of similar magnitude than for graphene based biosensors (see Fig. 9.12). The conventional PB model clearly overestimates the capacitance of (100) oriented hydrogen-terminated diamond biosensors because it does not take into account the depletion of water and ions. The hydrophilic model, however, takes into account PMFs for the ions and a spatially varying water density. It is interesting to see that both the PB and the extended PB model for a hydrophilic interface lead to similar capacitances (around 7 $\mu\mathrm{F/cm^2}$, not shown). In the latter case the PMFs and the water density profile for a hydrophilic interface were used rather than the ones for a hydrophobic interface. In the hydrophilic case, both the ions and the water can approach the surface closer but still not as close and not with such a high density as with the PB model. The shortcomings of the PB model are less pronounced for hydrophilic surfaces such as the Si–SiO_2–electrolyte interface, investigated in Chapter 7.

In conclusion we have demonstrated that the hole accumulation at the hydrogen-terminated diamond–electrolyte interface can therefore be simulated to great accuracy with the extended Poisson–Boltzmann model where the hydrophobic character of the surface is explicitly taken into consideration. Hydrophobicity limits the approach of electrolyte ions to the surface and therefore increases the potential drop in the electrolyte. This reduces the effective potential at the semiconductor surface. Consequently, the sensitivity of potentiometric biosensor devices which depend on the variation of the number of charge carriers with potential change is profoundly affected by the modification of the interfacial capacitance for any hydrophobic surface.

A. Metal–insulator–semiconductor structures

Based on the previous work of Ref. [Hac02], we are describing here in little more detail the modeling of metal-insulator-semiconductor (MIS) contacts that are very important for MOSFET simulations. They are also termed MOS (metal–oxide–semiconductor) contacts. A very good review about MOS structures can be found in [Sze81]. In this work we are treating a MIS contact as a Schottky contact by specifying an appropriate Schottky barrier.

When a metal is in contact with a semiconductor, a potential barrier is formed at the metal–semiconductor interface. In 1938, Walter Schottky suggested that this potential barrier arises due to stable space charges in the semiconductor [Sch38]. At thermal equilibrium, the Fermi levels of the metal and the semiconductor must coincide. There are two limiting cases:

- **Ideal Schottky barrier**

  In a metal–*n*-type semiconductor structure, the barrier height ϕ_{B} is the difference of the metal work function ϕ_{m} and the electron affinity χ_{s} of the semiconductor

  $$e\phi_{\mathrm{B}} = e\left(\phi_{\mathrm{m}} - \chi_{\mathrm{s}}\right), \tag{A.1}$$

  where e is the positive elementary charge.
  In a metal–*p*-type semiconductor structure, the barrier height $\phi_{\mathrm{B,p}}$ is given by

  $$-e\phi_{\mathrm{B,p}} = e\left(\phi_{\mathrm{m}} - \chi_{\mathrm{s}}\right) - E_{\mathrm{gap}}, \tag{A.2}$$

  where E_{gap} is the band gap energy.

- **Fermi level pinning**

  If surface states on the semiconductor surface are present, then the barrier height ϕ_{B} is determined by the property of the semiconductor surface and is independent of the metal work function ϕ_{m}. The conduction band edge is then pinned at $e\phi_{\mathrm{B}}$ above the Fermi level.

The Schottky barrier model of next**nano**^{3} is implemented as a Fermi level pinning, where we assume that the conduction band edge E_{c} is pinned with respect to the Fermi level E_{F} due to surface states (interface states). Thus the barrier height is independent of the metal work function and is entirely determined by the surface states. Although it is not possible to automatically take into account the work function of the metal, we will

show in the following how the Schottky barrier height is related to the work functions of the metal and the semiconductor, and to the electron affinities of the insulator and the semiconductor. The barrier height can thus be adjusted manually to take into account the dependence on electron affinities, doping concentrations or surface charge. For a Schottky contact, only the barrier height ϕ_B and the applied voltage are needed as input parameters.

Figure A.1 shows the calculated conduction band edge profile of a Schottky contact at a metal–semiconductor interface at zero gate bias ($V_\mathrm{GS} = 0\,\mathrm{V}$). In this example, the semiconductor is GaAs with an n-type doping concentration of $10^{18}\,\mathrm{cm}^{-3}$ (fully ionized) at $T = 300\,\mathrm{K}$ and a Schottky barrier height of $\phi_\mathrm{B} = 0.53\,\mathrm{V}$ (solid line). If the semiconductor is doped, the conduction and valence band edges are shifted with respect to the Fermi level, and thus depend on doping. This is a bulk property and independent of surface effects, such as an ohmic contact or a Schottky barrier height (see right boundary of Fig. A.1). At the left boundary, however, the band profile is affected by the type of contact. For comparison, the dotted line shows the calculated conduction band profile for a Schottky barrier of $\phi_\mathrm{B} = 0\,\mathrm{V}$. The dashed line shows the conduction band edge profile for an ohmic contact where one assumes Neumann boundary conditions in the Poisson equation (eq. (1.1)), i.e. the derivate of the electrostatic potential $\frac{\partial\phi}{\partial x} = -F_x = 0$. The latter is the flat band condition, i.e. zero slope for the electrostatic potential, which is equivalent to a vanishing electric field $\mathbf{F}$. Note that a Schottky barrier of $\phi_\mathrm{B} = 0\,\mathrm{V}$ is not equivalent to an ohmic contact. A Schottky barrier is a Dirichlet boundary condition for the electrostatic potential, thus fixing the value of the conduction band edge at the surface with respect to the Fermi level. This is because the semiconductor band edge energies at the metal–semiconductor interface are in a definite energy relationship with the Fermi level of the metal

$$E_\mathrm{c} - E_\mathrm{F} = e\phi_\mathrm{B}. \tag{A.3}$$

In this particular example, an artificial Schottky barrier of $\phi_\mathrm{B} = -0.042\,\mathrm{V}$ would be equivalent to an ohmic contact, but only for the same temperature and the same doping concentration.

Instead of specifying a Schottky barrier ϕ_B, one can alternatively specify an interface charge density of the surface states at the metal–semiconductor interface. A fixed interface sheet charge density of $\sigma = -4.434 \cdot 10^{-3}\,\mathrm{As/m}^2$ (corresponding to a sheet charge carrier concentration of $-2.768 \cdot 10^{12}\,\mathrm{cm}^{-2}$) leads to exactly the same conduction band edge profile (not shown) as for $\phi_\mathrm{B} = 0.53\,\mathrm{V}$ in Fig. A.1. In this case, the interface charge density corresponds to a Neumann boundary condition for the electrostatic potential with a fixed, nonzero slope, i.e. a nonzero electric field at the boundary ($\frac{\partial\phi}{\partial x} = -F_x \neq 0$). The electric field F_x is related to the sheet charge density

$$F_x = \frac{\sigma}{\varepsilon_0\varepsilon_\mathrm{r}}, \tag{A.4}$$

where ε_0 is the vacuum permittivity and ε_r is the static dielectric constant of the semiconductor. In this example ($\varepsilon_\mathrm{r} = 12.93$ for GaAs), the electric field at the surface is

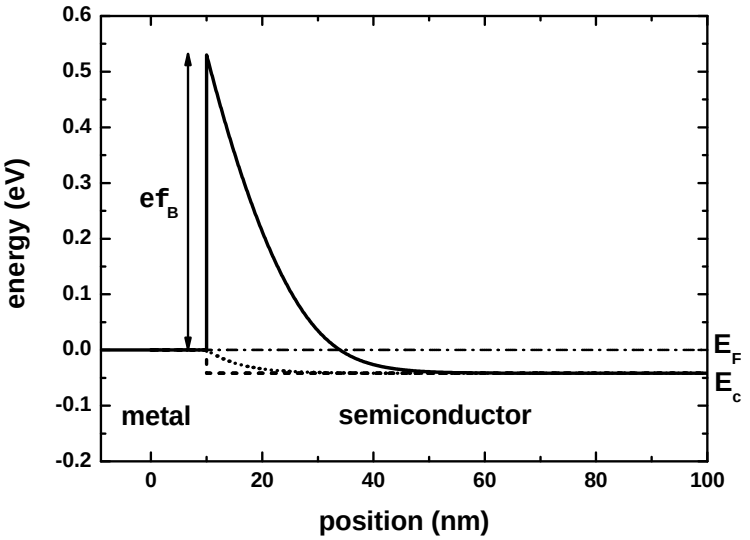


Figure A.1.: Calculated conduction band profile of a Schottky contact at a metal–semiconductor interface (n-type GaAs). The Schottky contact is completely specified by the Schottky barrier height $\phi_{\mathrm{B}} = 0.53\,\mathrm{V}$ (solid line). The conduction band edge E_{c} is pinned with respect to the Fermi level $E_{\mathrm{F}} = 0\,\mathrm{eV}$ (dash-dotted line) due to surface states. For comparison, the dotted line shows the calculated conduction band profile for $\phi_{\mathrm{B}} = 0\,\mathrm{V}$. The dashed line shows the conduction band profile for an ohmic contact where one assumes a flat band boundary condition for the electrostatic potential.

387 kV/cm. Numerically, within the next**nano**[3] software the interface sheet charge density is converted into a volume charge density and enters the Poisson equation as a fixed charge density (eq. (1.2)).

In the engineering literature for MOS structures two values are important. The work function of the metal ϕ_{m} and the work function of the semiconductor ϕ_{s}. The work function ϕ_{m} is the energy that is needed to move an electron at the Fermi level $E_{\mathrm{F,m}}$ of the metal to the vacuum level E_{vac}

$$e\phi_{\mathrm{m}} = E_{\mathrm{vac}} - E_{\mathrm{F,m}}. \tag{A.5}$$

For the semiconductor the work function reads analogously

$$e\phi_{\mathrm{s}} = E_{\mathrm{vac}} - E_{\mathrm{F,s}}, \tag{A.6}$$

where the work function depends on the doping properties. Another quantity that has to be defined is the electron affinity which describes the energy difference between the vacuum level and the conduction band edge. The electron affinity of the insulator is

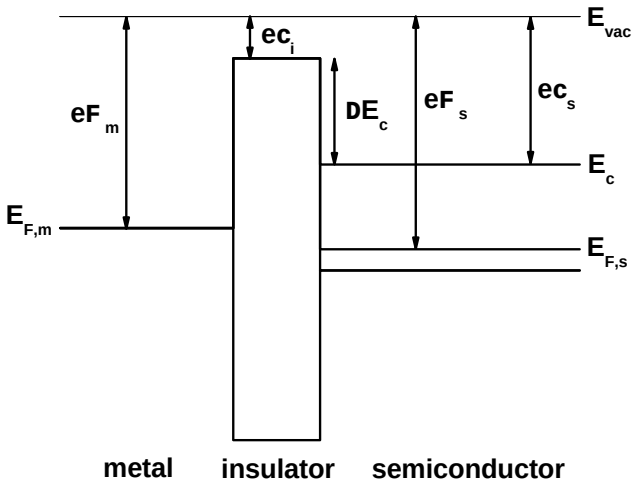


Figure A.2.: Band diagram of a MIS structure which is not in thermal equilibrium. ϕ_m is the work function of the metal, ϕ_s of the semiconductor, χ_i is the electron affinity of the insulator, χ_s of the semiconductor, $E_\mathrm{F,m}$ is the Fermi level of the metal, $E_\mathrm{F,s}$ of the semiconductor, E_c is the conduction band edge of the semiconductor and ΔE_c is the conduction band discontinuity at the interface between the insulator and the semiconductor.

labeled by χ_i, the one of the semiconductor by χ_s. These quantities are depicted in Fig. A.2 for a MIS structure that is not in thermal equilibrium. Once the MIS structure is in thermal equilibrium, space charges are created that lead to a built-in potential. Thus the vacuum energy is shifted accordingly, resulting in a constant Fermi level of the whole device. This is shown in Fig. A.3 where the Poisson equation has been solved in the insulator–semiconductor region of a metal–SiO_2–Si structure at $T = 300\,\mathrm{K}$. The silicon layer is p-type doped with boron with a concentration of $3 \cdot 10^{17}\,\mathrm{cm}^{-3}$. For the Schottky contact the barrier height ϕ_B reads

$$\phi_\mathrm{B} = \phi_\mathrm{m} - \chi_\mathrm{i}. \tag{A.7}$$

Furthermore it holds for the conduction band discontinuity ΔE_c at the interface between the insulator and the semiconductor

$$\Delta E_\mathrm{c} + e\chi_\mathrm{i} = e\chi_\mathrm{s}. \tag{A.8}$$

Thus it follows

$$e\phi_\mathrm{B} = e\phi_\mathrm{m} - e\chi_\mathrm{s} + \Delta E_\mathrm{c}, \tag{A.9}$$

which can be written in terms of the difference of the work functions

$$e\phi_\mathrm{B} = e\,(\phi_\mathrm{m} - \phi_\mathrm{s}) + (E_\mathrm{c} - E_\mathrm{F,s}) + \Delta E_\mathrm{c}. \tag{A.10}$$

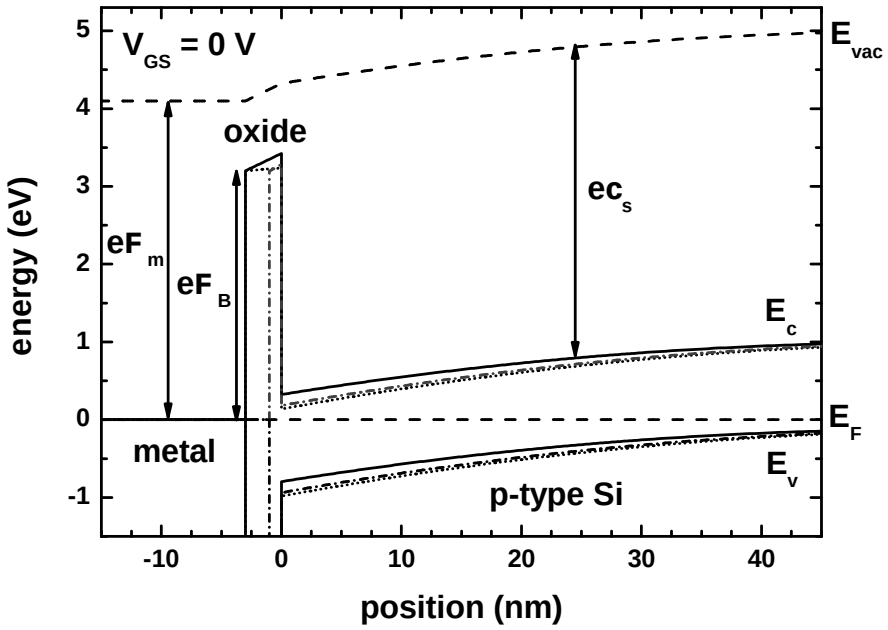


Figure A.3.: Band diagram of a MOS structure in thermal equilibrium. Technically, one can assume a Schottky barrier of height $e\phi_{\mathrm{B}} = e\phi_{\mathrm{m}} - e\chi_{\mathrm{s}} + \Delta E_{\mathrm{c}}$ as the boundary condition. Due to the built-in potential the vacuum energy is bent (only shown for the case of the 3 nm SiO_2 layer (solid line)). For comparison, the conduction and valence band edges for a 1 nm SiO_2 gate dielectric (dash-dotted line), and for a 3 nm high-k gate dielectric (HfO_2, $\varepsilon_{\mathrm{r}} = 26.1$, dotted line) are included.

In this equation both ϕ_{s} and $E_{\mathrm{c}} - E_{\mathrm{F,s}}$ depend on the doping of the semiconductor, whereas in eq. (A.9) only doping independent material parameters are used. Thus it is sufficient to know the work function ϕ_{m} of the metal, the electron affinity χ_{s} of the semiconductor and the conduction band offset ΔE_{c} between the insulator and the semiconductor in order to determine the appropriate Schottky barrier boundary condition.

In Fig. A.3, a Schottky barrier height of $\phi_{\mathrm{B}} = 3.2\,\mathrm{V}$ has been used. The conduction and valence band edges for a 1 nm SiO_2 gate dielectric (dash-dotted line), and for a 3 nm high-k gate dielectric (HfO_2, $\varepsilon_{\mathrm{r}} = 26.1$, dotted line) are shown for comparison. Reducing the oxide thickness or increasing the dielectric constant of the oxide has a similar effect on the conduction band edge profile, and thus on the charge carrier density. For simplicity, and in order to better demonstrate the influence of the large dielectric constant of HfO_2, all other parameters were taken to be the ones of SiO_2. Nowadays, a lot of effort is put into the optimization of metal–SiO_2–Si or metal–HfO_2–Si contacts in order to continue the down-scaling of CMOS transistors.

B. Temperature dependent material parameters

I have implemented temperature dependent material parameters into the next**nano**^{3} software only if reasonable interpolation formulas were available. This is currently the case for lattice constants, band gaps and band gap dependent $\mathbf{k} \cdot \mathbf{p}$ parameters.

B.1. Temperature dependent lattice constants

The lattice constants that are given in the database are only relevant for the strain calculations, although there are some exceptions to this rule for very specialized features like the calculation of alloy scattering. They vary with temperature and corresponding coefficients b (in units of nm/K) for the most important semiconductor materials are available. The lattice constant in the database is given for 300 K. For all other temperatures T (in units of K), the lattice constant a is calculated by

$$a = a^{300\,\mathrm{K}} + b(T - 300\,\mathrm{K}). \tag{B.1}$$

The parameters for the temperature dependent lattice constants in the database were taken from Ref. [VMRM01]. For cubic crystals, the lattice constant is specified by a, in wurtzite two lattice constants are required, namely a in the plane perpendicular to the [0001] axis, and c parallel to the [0001] axis. Unfortunately, no expansion coefficients for nitrides are given in Ref. [VMRM01] and Ref. [VM03]. For these materials more complicated formulas are necessary. They take into account that, for instance, AlN has a negative expansion coefficient below 100 K.

B.2. Temperature dependent band gaps

In a bulk semiconductor, both direct (Γ) and indirect band gaps (L, X) depend on temperature. The variation between 0 K and 300 K is of the order 0.1 eV and must be considered in realistic simulations. Most common is the empirical Varshni formula where the functional form is fitted to

$$E^i_{\mathrm{gap}}(T) = E^{i,0\,\mathrm{K}}_{\mathrm{gap}} - \frac{\alpha^i T^2}{T + \beta^i}, \tag{B.2}$$

where α and β are the Varshni parameters given in the database for each material and for each band gap ($i = \Gamma, \mathrm{L}, \mathrm{X}$). In the literature several different combinations of Varshni parameters exist for each material depending on which temperature range

should be fitted most accurately. The Varshni parameters in the database were taken from Ref. [VMRM01].

Experimentally, one only knows the temperature dependence of the band gap but unfortunately not the corresponding individual shifts of the valence and conduction band edge energies. Within the next**nano**[3] program, the valence and conduction band edges are defined for all materials on a global scale [VdW89]. For simplicity, we assume that for all materials the valence band edges are independent of temperature. Consequently, also the valence band offsets are independent of temperature. It follows that the temperature dependence of the band gap only affects the conduction band edge energies $E_{\mathrm{c}}^{i}(T)$. This implies that only the conduction band offsets are temperature dependent. However, this is not entirely correct as in reality also the valence band offsets depend on temperature, although not as much as the conduction band offsets. Optionally, one can manually adjust the conduction and valence band offsets if more reliable experimental data is available.

For a ternary alloy $\mathrm{A}_x\mathrm{B}_{1-x}$ composed of two binary materials A and B, the Varshni parameters are not interpolated linearly. Instead, the following procedure is applied. First, the average valence band edge energy $E_{\mathrm{v,av}}$ (see Fig. 3.1) of the ternary is calculated according to eq. (B.3) taking into account a possible bowing of the valence band offset. Then the *temperature independent* valence band edges of the ternary are calculated using the split-off energy Δ_{so} of the ternary, where a possible bowing of Δ_{so} has been taken care of. Then for each binary, the temperature dependent band gaps at Γ, L and X are calculated using the respective Varshni parameters. The next step is to calculate the band gaps of the ternary from the binary band gaps including bowing. Finally, the *temperature dependent* conduction band edge energies of the ternary are obtained by adding the band gap energies of the ternary to the value of the topmost valence band edge energy of the ternary.

The equation for interpolating a material parameter M is given by

$$M^{\mathrm{A}_x\mathrm{B}_{1-x}} = xM^{\mathrm{A}} + (1-x)M^{\mathrm{B}} - x(1-x)C, \tag{B.3}$$

where $M^{\mathrm{A}_x\mathrm{B}_{1-x}}$ is the resulting material parameter of the ternary alloy and x is the alloy composition. C is the bowing parameter. It accounts for deviations of the linear interpolation (virtual-crystal approximation) by introducing a quadratic term. Occasionally, the last term is written with the opposite sign. Then the sign of C has to be adjusted accordingly. The linear interpolation of the lattice constants is called Vegard's law. Interpolation formulas for quaternary materials are discussed in Ref. [Zib07] and Ref. [VMRM01].

B.3. Temperature dependent $\mathbf{k} \cdot \mathbf{p}$ parameters

The following 8×8 $\mathbf{k} \cdot \mathbf{p}$ parameters are temperature dependent because they depend on the band gap E_{gap} at the Γ point.

Zinc blende

- S (eq. (3.62))
- the modified DKK parameters L', $N^{+\prime}$, N' (eq. (3.101), eq. (3.103), eq. (3.105))
- the modified Luttinger parameters γ_1', γ_2', γ_3', κ' (eq. (3.118), eq. (3.119), eq. (3.120), eq. (3.121))
- the modified Foreman parameter σ' (eq. (3.126))
- F' (eq. (3.122))
- If the $\mathbf{k} \cdot \mathbf{p}$ parameters are rescaled, also the Kane parameter E_{P} (eq. (3.158) or eq. (3.159), respectively) is affected, and consequently also the Kane momentum matrix element P (eq. (3.79)).

Wurtzite

- S_1, S_2 (eq. (3.63), eq. (3.64))
- the modified DKK parameters L_1', L_2', $N_1^{+\prime}$, $N_2^{+\prime}$, N_1', N_2' (eq. (3.106), eq. (3.107), eq. (3.108), eq. (3.109), eq. (3.110), eq. (3.111))
- the modified RSP parameters A_1', A_3', A_4', A_5', A_6' (eq. (3.112), eq. (3.114), eq. (3.115), eq. (3.116), eq. (3.117))
- the modified Foreman parameters σ', σ_z', σ_{xz}' (eq. (3.42))
- If the $\mathbf{k}\cdot\mathbf{p}$ parameters are rescaled, also the Kane parameters E_{P1} and E_{P2} (eq. (3.63), eq. (3.64)) are affected, and consequently also the Kane momentum matrix elements P_1 and P_2 (eq. (3.81), eq. (3.82)).

C. Analytical equations for biaxial strain for arbitrary substrate orientations

Here, we list analytical equations for the strain tensor of cubic crystals. They are valid for heterostructures that are homogeneous along two directions, and for arbitrary substrate orientations. First, we present the results of our derivation of the equations for the simulation coordinate system [PGBD+11], then we list the results for the crystal coordinate system.

C.1. Simulation coordinate system

The strain tensor components ϵ'_{ij} in the simulation coordinate system basis $\mathbf{a} = (a, b, c)$ for arbitrary substrate orientation indicated by the Miller index (hkl) are given by

$$\epsilon' = \begin{pmatrix} 1 & 0 & D_1^M \\ 0 & 1 & D_2^M \\ D_1^M & D_2^M & -D_0^M \end{pmatrix} \epsilon'_{\parallel}, \tag{C.1}$$

where $D_0^M = -\epsilon'_{cc}/\epsilon'_{aa} = -\epsilon'_{\perp}/\epsilon'_{\parallel}$ is the biaxial Poisson ratio (notated D^M by Van de Walle [VdW89]). Here, the substrate is oriented parallel to the (a, b) plane, and the growth direction is the c axis perpendicular to it. We define $D_1^M = \epsilon'_{ca}/\epsilon'_{aa}$ and $D_2^M = \epsilon'_{bc}/\epsilon'_{aa}$ as the shear-to-biaxial strain ratios. Finally, the strain tensor components read

$$\epsilon'_{aa} = \epsilon'_{bb} = \epsilon'_{\parallel} = \frac{a_{\text{substrate}} - a_{\text{layer}}}{a_{\text{layer}}} \tag{C.2}$$

$$\epsilon'_{ab} = \epsilon'_{ba} = 0 \tag{C.3}$$

$$\epsilon'_{ac} = \epsilon'_{ca} = \frac{\lambda\mu - \eta\omega}{\lambda\kappa - \eta^2}\epsilon'_{\parallel} = D_1^M \epsilon'_{\parallel} \tag{C.4}$$

$$\epsilon'_{bc} = \epsilon'_{cb} = \frac{\omega - \eta D_1^M}{\lambda}\epsilon'_{\parallel} = D_2^M \epsilon'_{\parallel} \tag{C.5}$$

$$\epsilon'_{cc} = \epsilon'_{\perp} = \frac{\alpha - 2C'_{34}D_2^M - 2C'_{35}D_1^M}{C'_{33}}\epsilon'_{\parallel} = -D_0^M \epsilon'_{\parallel}, \tag{C.6}$$

where the denominators in eq. (C.4), eq. (C.5) and eq. (C.6) are always nonzero and

$$\alpha = -\left(C'_{13} + C'_{23}\right) \tag{C.7}$$
$$\beta = -\left(C'_{14} + C'_{24}\right) \tag{C.8}$$
$$\gamma = -\left(C'_{15} + C'_{25}\right) \tag{C.9}$$
$$\lambda = 2C'_{33}C'_{44} \tag{C.10}$$
$$\kappa = 2C'_{33}C'_{55} \tag{C.11}$$
$$\eta = 2\left(C'_{33}C'_{45} - C'_{34}C'_{35}\right) \tag{C.12}$$
$$\omega = C'_{33}\beta - C'_{34}\alpha \tag{C.13}$$
$$\mu = C'_{33}\gamma - C'_{35}\alpha. \tag{C.14}$$

The coefficients α, β, and γ are first order, whereas the coefficients λ, κ, η, ω and μ are second order in the elastic stiffness tensor components C'_{ij}. The 6×6 matrix C'_{ij} contains the elastic constants in the Voigt notation with respect to the simulation coordinate system basis $\mathbf{a}$. It is obtained by rotating the forth-rank elastic stiffness tensor C_{ijkl} from the crystal coordinate system basis $\mathbf{x} = (x, y, z)$ to the simulation coordinate system basis $\mathbf{a}$

$$C'_{mnop} = \mathbf{R}^M_{mi}\mathbf{R}^M_{nj}\mathbf{R}^M_{ok}\mathbf{R}^M_{pl}C_{ijkl}, \tag{C.15}$$

where $\mathbf{R}^M$ is the rotation matrix defined as

$$\mathbf{a} = \mathbf{R}^M\mathbf{x}. \tag{C.16}$$

The rotated forth-rank tensor C'_{mnop} has to be mapped into the contracted C'_{ij} Voigt notation that has more nonzero entries as compared to C_{ij}. The strain tensor ϵ' can be rotated into the crystal coordinate system basis using

$$\epsilon = \left(\mathbf{R}^M\right)^{-1}\epsilon'\mathbf{R}^M. \tag{C.17}$$

C.2. Crystal coordinate system

The strain tensor components ϵ_{ij} in the crystal coordinate system $\mathbf{x} = (x, y, z)$ for arbitrary substrate orientation (hkl) are obtained by symmetrizing the components $\tilde{u}_{ij}$ of the distortion tensor (see eq. (1.8)). The distortion tensor $\tilde{\mathbf{u}}$ is identical to the strain tensor for high symmetry orientations like (001), (110) and (111) but in general it is nonsymmetric for low symmetry substrate orientations and given by [JaM, And09]

$$\tilde{\mathbf{u}} = \begin{pmatrix} u_0 & 0 & 0 \\ 0 & u_0 & 0 \\ 0 & 0 & u_0 \end{pmatrix} + D \begin{pmatrix} n_1^2 D_1 & n_1 n_2 D_1 & n_1 n_3 D_1 \\ n_2 n_1 D_2 & n_2^2 D_2 & n_2 n_3 D_2 \\ n_3 n_1 D_3 & n_3 n_2 D_3 & n_3^2 D_3 \end{pmatrix}, \tag{C.18}$$

where n_1, n_2 and n_3 are the components of the normalized vector $\hat{\mathbf{c}}$ along the growth direction $[hkl]$

$$\hat{\mathbf{c}} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \frac{1}{\sqrt{h^2 + k^2 + l^2}} \begin{pmatrix} h \\ k \\ l \end{pmatrix}. \tag{C.19}$$

The lattice mismatch is defined as

$$u_0 = \frac{a_{\text{substrate}} - a_{\text{layer}}}{a_{\text{layer}}}, \tag{C.20}$$

and the other variables are given by

$$D = \frac{-u_0 \left(C_{11} + 2C_{12}\right)}{D_x D_y D_z + \left(C_{12} + C_{44}\right)\left(n_1^2 D_1 + n_2^2 D_2 + n_3^2 D_3\right)} \tag{C.21}$$

$$D_1 = D_y D_z \tag{C.22}$$

$$D_2 = D_z D_x \tag{C.23}$$

$$D_3 = D_x D_y \tag{C.24}$$

$$D_x = \left(C_{11} - C_{12} - C_{44}\right) n_1^2 + C_{44}\left(n_2^2 + n_3^2\right) \tag{C.25}$$

$$D_y = \left(C_{11} - C_{12} - C_{44}\right) n_2^2 + C_{44}\left(n_3^2 + n_1^2\right) \tag{C.26}$$

$$D_z = \left(C_{11} - C_{12} - C_{44}\right) n_3^2 + C_{44}\left(n_1^2 + n_2^2\right). \tag{C.27}$$

To rotate the strain tensor ϵ into the simulation coordinate system basis, use

$$\epsilon' = \mathbf{R}^M \epsilon \left(\mathbf{R}^M\right)^{-1}. \tag{C.28}$$

List of Figures

List of Tables

List of Publications

1. *Hydrophobic interaction and charge accumulation at the diamond/electrolyte interface*
   M. Dankerl, A. Lippert, S. Birner, E. U. Stützel, M. Stutzmann, J. A. Garrido
   Physical Review Letters **106**, 196103 (2011)

2. *Valley degeneracy in biaxially strained aluminum arsenide quantum wells*
   S. Prabhu-Gaunkar, S. Birner, S. Dasgupta, C. Knaak, M. Grayson
   Physical Review B **84**, 125319 (2011)

3. *Electron heating in quantum-dot structures with collective potential barriers*
   L. H. Chien, A. Sergeev, N. Vagidov, V. Mitin, S. Birner
   International Journal of High Speed Electronics and Systems **20**, 143 (2011)

4. *Graphene Solution-Gated Field-Effect Transistor Array for Sensing Applications*
   M. Dankerl, M. Hauf, A. Lippert, L. Hess, S. Birner, I. D. Sharp, A. Mahmood, P. Mallet, J. Y. Veuillen, M. Stutzmann, J. A. Garrido
   Advanced Functional Materials **20**, 3117 (2010)

5. *Electronic and optical properties of GaN/AlN quantum dots on Si(111) subject to in-plane uniaxial stresses and variable excitation*
   O. Moshe, D. H. Rich, S. Birner, M. Povolotskyi, B. Damilano, J. Massies
   Journal of Applied Physics **108**, 083510 (2010)

6. *Ballistic quantum transport using the contact block reduction (CBR) method – An introduction*
   S. Birner, C. Schindler, P. Greck, M. Sabathil, P. Vogl
   Journal of Computational Electronics **8**, 267–286 (2009)

7. *GaN/AlN short-period superlattices for intersubband optoelectronics: A systematic study of their epitaxial growth, design, and performance*
   P. K. Kandaswamy, F. Guillot, E. Bellet-Amalric, E. Monroy, L. Nevou, M. Tchernycheva, A. Michon, F. H. Julien, E. Baumann, F. R. Giorgetta, D. Hofstetter, T. Remmele, M. Albrecht, S. Birner, Le Si Dang
   Journal of Applied Physics **104**, 093501 (2008)

8. *Single-valley high-mobility (110) AlAs quantum wells with anisotropic mass*
   S. Dasgupta, S. Birner, C. Knaak, M. Bichler, A. Fontcuberta i Morral, G. Abstreiter, M. Grayson
   Applied Physics Letters **93**, 132102 (2008)

9. *Theoretical model for the detection of charged proteins with a silicon-on-insulator sensor*
   S. Birner, C. Uhl, M. Bayer, P. Vogl
   Journal of Physics: Conference Series **107**, 012002 (2008)

10. *Simulation of quantum cascade lasers – optimizing laser performance*
    S. Birner, T. Kubis, P. Vogl
    Photonik international **2**, 60 (2008)

11. *Simulation zur Optimierung von Quantenkaskadenlasern*
    S. Birner, T. Kubis, P. Vogl
    Photonik **1**, 44 (2008)

12. *Bandstructure and photoluminescence of SiGe islands with controlled Ge concentration*
    M. Brehm, T. Suzuki, Z. Zhong, T. Fromherz, J. Stangl, G. Hesser, S. Birner, F. Schäffler, G. Bauer
    Microelectronics Journal **39**, 485 (2008)

13. *nextnano: General Purpose 3-D Simulations*
    S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, P. Vogl
    IEEE Transactions on Electron Devices **54**, 2137 (2007)

14. *GaN quantum dots as optical transducers in field-effect chemical sensors*
    O. Weidemann, G. Jegert, S. Birner, M. Stutzmann, M. Eickhoff, E. Monroy
    2007 IEEE SENSORS **1**–**3**, 1175 (2007)
    Proceedings of the 6$^{\text{th}}$ IEEE Sensors Conference, Atlanta, GA

15. *Three-Dimensional Si/Ge Quantum Dot Crystals*
    D. Grützmacher, T. Fromherz, C. Dais, J. Stangl, E. Müller, Y. Ekinci, H. H. Solak, H. Sigg, R. T. Lechner, E. Wintersberger, S. Birner, V. Holý, G. Bauer
    Nano Letters **7** (10), 3150 (2007)

16. *Two-photon excitation of asymmetric GaN/AlGaN quantum discs*
    K. H. Lee, J. H. Na, S. Birner, S. N. Yi, R. A. Taylor, Y. S. Park, C. M. Park, T. W. Kang
    AIP Conf. Proc. **893**, 981 (2007)
    Proceedings of the 28$^{\text{th}}$ International Conference on the Physics of Semiconductors, Vienna (2006)

17. *Characterization of tunneling and free-carrier screening in coupled asymmetric GaN/AlGaN quantum discs*
    K. H. Lee, J. H. Na, S. Birner, S. N. Yi, R. A. Taylor, Y. S. Park, C. M. Park, T. W. Kang
    AIP Conf. Proc. **893**, 1003 (2007)
    Proceedings of the 28$^{\text{th}}$ International Conference on the Physics of Semiconductors, Vienna (2006)

18. *The 3D nanometer device project nextnano: Concepts, methods, results*
A. Trellakis, T. Zibold, T. Andlauer, S. Birner, R. K. Smith, R. Morschl, P. Vogl
Journal of Computational Electronics **5**, 285 (2006)

19. *Two-photon excitation spectroscopy of coupled asymmetric GaN/AlGaN quantum discs*
K. H. Lee, S. Birner, J. H. Na, R. A. Taylor, S. N. Yi, Y. S. Park, C. M. Park, T. W. Kang
Nanotechnology **17**, 5754 (2006)

20. *Modeling of semiconductor nanostructures with next**nano**^{3}*
S. Birner, S. Hackenbuchner, M. Sabathil, G. Zandler, J. A. Majewski, T. Andlauer, T. Zibold, R. Morschl, A. Trellakis, P. Vogl
Acta Physica Polonica A **110** (2), 111 (2006)

21. *Enhancement of free-carrier screening due to tunneling in coupled asymmetric GaN/AlGaN quantum discs*
K. H. Lee, J. H. Na, R. A. Taylor, S. N Yi, S. Birner, Y. S. Park, C. M. Park, T. W. Kang
Applied Physics Letters **89**, 023103 (2006)

22. *Free carrier screening in coupled asymmetric GaN quantum discs*
K. H. Lee, J. H. Na, S. Birner, R. A. Taylor, S. N Yi, Y. S. Park, C. M. Park, T. W. Kang
SPIE Proceedings **6352**, U288–U296 (2006)
Proceedings of the Conference on Optoelectronic Materials and Devices, Gwangju, South Korea

23. *Bound-to-bound and bound-to-free transitions in surface photovoltage spectra: Determination of the band offsets for $In_xGa_{1-x}As$ and $In_xGa_{1-x}As_{1-y}N_y$ quantum wells*
M. Galluppi, L. Geelhaar, H. Riechert, M. Hetterich, A. Grau, S. Birner, W. Stolz
Physical Review B **72**, 155324 (2005)

24. *Effects of strain and confinement on the emission wavelength of InAs quantum dots due to a $GaAs_{1-x}N_x$ capping layer*
O. Schumann, S. Birner, M. Baudach, L. Geelhaar, H. Eisele, L. Ivanova, R. Timm, A. Lenz, S. K. Becker, M. Povolotskyi, M. Dähne, G. Abstreiter, H. Riechert
Virtual Journal of Nanoscale Science Technology **12** (1) (2005) &
Physical Review B **71**, 245316 (2005)

25. *Modeling the nonlinear photoluminescence intensity dependence observed in asymmetric GaN quantum discs with AlGaN barriers*
K. H. Lee, S. Birner, J. H. Na, R. A. Taylor, J. W. Robinson, J. H. Rice, Y. S. Park, C. M. Park, T. W. Kang
Proceedings of 2005 5$^{\text{th}}$ IEEE Conference on Nanotechnology, Nagoya, Japan, 547 (2005)

26. *Modeling of Purely Strain-Induced CEO GaAs/$In_{0.16}Al_{0.84}As$ Quantum Wires*
S. Birner, R. Schuster, M. Povolotskyi, P. Vogl
Proceedings of the 5$^{\text{th}}$ International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD'05), Berlin, Germany, 1 (2005)

27. *Purely strain induced GaAs/InAlAs single quantum wires exhibiting strong charge carrier confinement*
R. Schuster, H. Hajak, M. Reinwald, W. Wegscheider, D. Schuh, M. Bichler, S. Birner, P. Vogl, G. Abstreiter
AIP Conf. Proc. **772**, 898 (2005)
Proceedings of the 27$^{\text{th}}$ International Conference on the Physics of Semiconductors, Flagstaff (2004)

28. *Optical properties of low-dimensional semiconductor systems fabricated by cleaved edge overgrowth*
R. Schuster, H. Hajak, M. Reinwald, W. Wegscheider, G. Schedelbeck, S. Sedlmaier, M. Stopa, S. Birner, P. Vogl, J. Bauer, D. Schuh, M. Bichler, G. Abstreiter
physica status solidi (c) **1**, 2028 (2004)

29. *Advances in the theory of electronic structure of semiconductors*
J. A. Majewski, S. Birner, A. Trellakis, M. Sabathil, P. Vogl
physica status solidi (c) **1**, 2003 (2004)

30. *Electronic and optical properties of [N11] grown nanostructures*
M. Povolotskyi, A. Di Carlo, S. Birner
physica status solidi (c) **1**, 1511 (2004)

31. *Tuning the Piezoelectric Fields in Quantum Dots: Microscopic Description of Dots Grown on (N11) Surfaces*
M. Povolotskyi, A. Di Carlo, P. Lugli, S. Birner, P. Vogl
IEEE Transactions on Nanotechnology **3**, 124 (2004)

32. *Non-linear optical properties of InGaAs/AlGaAs nanostructures grown on (N11) surfaces*
M. Povolotskyi, J. Gleize, A. Di Carlo, P. Lugli, S. Birner, P. Vogl, D. Alderighi, M. Gurioli, A. Vinattieri, M. Colocci, S. Sanguinetti, R. Nötzel
Semiconductor Science and Technology **19**, S351 (2004)

33. *Two-dimensional quantum-mechanical modeling for strained silicon channels of double-gate MOSFETs*
K. Kim, O. Kwon, J. Seo, T. Won, S. Birner, R. Oberhuber, A. Trellakis
Journal of the Korean Physical Society **45**, S909–S913 (2004)

34. *Two-dimensional quantum effects and structural optimization of FinFETs with two-dimensional Poisson–Schrödinger solvers*
K. Kim, O. Kwon, J. Seo, T. Won, S. Birner, A. Trellakis
Journal of the Korean Physical Society **45**, 1384–1390 (2004)

35. *next**nano**^{3}– Software für neuartige Nano Devices*
S. Birner
TUM Mitteilungen **2** – 03/04, Technische Universität München, 36 (2003)

36. *Efficient computational method for ballistic currents and application to single quantum dots*
M. Sabathil, S. Birner, D. Mamaluy, P. Vogl
Journal of Computational Electronics **2**, 269 (2003)

37. *Microscopic description of nanostructures grown on (N11) surfaces*
M. Povolotskyi, J. Gleize, A. Di Carlo, P. Lugli, S. Birner, P. Vogl
Journal of Computational Electronics **2**, 275 (2003)

38. *Theory of vertical and lateral Stark shifts of excitons in InGaAs quantum dots*
M. Sabathil, S. Hackenbuchner, S. Birner, J. A. Majewski, P. Vogl, J. J. Finley
physica status solidi (c) **0**, 1181 (2003)

39. *Accelerated dynamics simulations of interstitial-cluster growth*
S. Birner, J. Kim, D. A. Richie, J. W. Wilkins, A. F. Voter, T. Lenosky
Solid State Communications **120**, 279 (2001)

40. *Thermodynamic processes of Si-interstitial clusters*
J. Kim, S. Birner, D. A. Richie, J. W. Wilkins, A. F. Voter
Computational Nanoscience **2001**, 117 (2001)

41. *Theory of interstitial aggregates in germanium*
S. Birner, J. P. Goss, R. Jones, P. R. Briddon, S. Öberg
Proceedings of the 2nd ENDEASD (European Network on Defect Engineering of Advanced Semiconductor Devices) Workshop, Kista, Sweden (2000)

Acknowledgements

First of all, I have to thank Prof. Peter Vogl for the opportunity to work on the exciting and very challenging field of developing a software for semiconductor nanostructures. He offered me the freedom to work on several interesting topics with always up-to-date hard- and software. The pleasant atmosphere at Prof. Vogl's chair is only possible thanks to the cooperativeness and collegiality of all my former and present colleagues: Prof. Jacek Majewski, Dr. Stefan Hackenbuchner, Dr. Günther Zandler, Dr. Matthias Sabathil, Dr. Alexandros Trellakis, Dr. Denis Mamaluy, Dr. Reinhard Scholz, Dr. Christian Strahberger, Dr. Tobias Zibold, Dr. Till Andlauer, Dr. Tillmann Kubis, Christoph Schindler, Peter Greck, Thomas Eißfeller, Michael Bayer, Christian Ammann, Christian Uhl, Philip Weidmann and Stefan Beer. I also greatly appreciated the intensive and illuminating discussions with the members of the other research groups at the Walter Schottky Institute (WSI). Special thanks are indebted to secretaries Veronika Enter, Joana Figueiredo and Liane Lindner.

I also have to thank Dr. Michael Povolotskyi, Prof. Aldo Di Carlo, Prof. Paolo Lugli, Prof. Vladimir Mitin, Dr. Nizami Vagidov, Prof. Martin Eickhoff, Prof. Martin Stutzmann, Prof. Matthew Grayson, Dr. Jose Garrido, Andreas Lippert, Markus Dankerl, Dr. Thomas Fromherz, Dr. Kwan H. Lee, Prof. Werner Wegscheider, Prof. Jan von Delft, Prof. Henning Riechert, Dr. Eva Monroy, Dr. Robert Schuster, Dr. Andrei Schliwa, Prof. Roderick Melnik, Prof. Eoin O'Reilly, Prof. Bernd Witzigmann, Prof. Joachim Piprek, Prof. Daniel Rich, Prof. Douglas Paul, Dr. Andreas Gehring, Dr. Anders Blom, Dr. Vittorio Zecca and the Nanosystems Initiative Munich for their collaborations and support.

During the last years, I had the opportunity to enjoy several fruitful collaborations worldwide with experimental physics and electrical engineering research groups, and also industry. Thus, I also want to thank all of them for their great work and all the fantastic things I have learned while working with them.

Finally, I also want to thank Christoph Zinser (Spin-off Service, LMU) and Dr. Bernward Jopen (UnternehmerTUM) for their help and valuable insights on the business aspects of the next**nano** software. Dr. Jörg Ehehalt did a great job for writing the graphical user interface to the next**nano** software. Thank you!

Most of all, however, I have to thank my wife, children, parents and parents-in-law for their patience and support during the long period of working on my thesis.

I really had a great time at the WSI! Thank you all.

Bibliography

[ACWL08] P. K. Ang, W. Chen, A. T. S. Wee, and K. P. Loh, *Solution-Gated Epitaxial Graphene as pH Sensor*, Journal of the American Chemical Society **130** (2008), no. 44, 14392–14393.

[AMV08] T. Andlauer, R. Morschl, and P. Vogl, *Gauge-invariant discretization in multiband envelope function theory and g factors in nanowire dots*, Physical Review B **78** (2008), no. 7, 075317.

[And04] T. Andlauer, *Discretization of multiband-$k \cdot p$-Schrödinger equations for multidimensional semiconductor nanostructures*, Master's thesis, Walter Schottky Institut and Physics Department, Technische Universität München, Germany, 2004.

[And09] ———, *Optoelectronic and spin-related properties of semiconductor nanostructures in magnetic fields*, Selected Topics of Semiconductor Physics and Technology (G. Abstreiter, M.-C. Amann, M. Stutzmann, and P. Vogl, eds.), vol. 105, Verein zur Förderung des Walter Schottky Instituts der Technischen Universität München e.V., München, 2009.

[Ant93] H. M. Antia, *Rational Function Approximations for Fermi–Dirac Integrals*, Astrophysical Journal Supplement **84** (1993), 101–108.

[AO00] A. D. Andreev and E. P. O'Reilly, *Theory of the electronic structure of GaN/AlN hexagonal quantum dots*, Physical Review B **62** (2000), no. 23, 15851–15870.

[AV09] T. Andlauer and P. Vogl, *Full-band envelope-function approach for type-II broken-gap superlattices*, Physical Review B **80** (2009), no. 3, 035304.

[Bah90] T. B. Bahder, *Eight-band $k \cdot p$ model of strained zinc-blende crystals*, Physical Review B **41** (1990), no. 17, 11992–12001.

[Bal10] K. Balasubramanian, *Challenges in the use of 1D nanostructures for on-chip biosensing and diagnostics: A review*, Biosensors and Bioelectronics **26** (2010), no. 4, 1195–1204.

[Bas88] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures*, Les Éditions de Physique, Les Ulis Cedex, France, 1988.

[BE96] R. J. Beynon and J. S. Easterby, *Buffer Solutions: The basics*, Oxford University Press, New York, 1996.

[Ber70] P. Bergveld, *Development of an ion-sensitive solid-state device for neurophysiological measurements*, IEEE Transactions on Biomedical Engineering **17** (1970), no. 1, 70–71.

[BHZ06] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, *Quantum Spin Hall Effect and Topological Phase Transitions in HgTe Quantum Wells*, Science **314** (2006), no. 5806, 1757–1761.

[BKN+05] H. Boroudjerdi, Y.-W. Kim, A. Naji, R. R. Netz, X. Schlagberger, and A. Serr, *Statics and dynamics of strongly charged soft matter*, Physics Reports **416** (2005), no. 3–4, 129–199.

[BKV08] S. Birner, T. Kubis, and P. Vogl, *Simulation of quantum cascade lasers – optimizing laser performance*, Photonik international **2** (2008), 60–63.

[BLA] BLAS – Basic Linear Algebra Subprograms, The BLAS source code is available from `http://www.netlib.org/blas`. Optimized libraries containing the BLAS routines are included in several compiler suites.

[Böe92] K. W. Böer, *Survey of Semiconductor Physics*, vol. II, Barriers, Junctions, Surfaces and Devices, Van Nostrand Reinhold, New York, 1992.

[Bon] D. J. Bonthuis, private communication.

[BP74] G. L. Bir and G. E. Pikus, *Symmetry and strain-induced effects in semiconductors*, John Wiley & Sons, New York, 1974.

[BRN+10] C. Brüne, A. Roth, E. G. Novik, M. König, H. Buhmann, E. M. Hankiewicz, W. Hanke, J. Sinova, and L. W. Molenkamp, *Evidence for the ballistic intrinsic spin Hall effect in HgTe nanostructures*, Nature Physics **6** (2010), no. 6, 448–454.

[BRRB+11] S. Boyer-Richard, F. Raouafi, A. Bondi, L. Pédesseau, C. Katan, J.-M. Jancu, and J. Even, *30-band $k \cdot p$ method for quantum semiconductor heterostructures*, Applied Physics Letters **98** (2011), no. 25, 251913.

[BSG+09] S. Birner, C. Schindler, P. Greck, M. Sabathil, and P. Vogl, *Ballistic quantum transport using the contact block reduction (CBR) method*, Journal of Computational Electronics **8** (2009), no. 3–4, 267–286.

[BUBV08] S. Birner, C. Uhl, M. Bayer, and P. Vogl, *Theoretical model for the detection of charged proteins with a silicon-on-insulator sensor*, Journal of Physics: Conference Series **107** (2008), 012002 (15pp).

[Bur92] M. G. Burt, *The justification for applying the effective-mass approximation to microstructures*, Journal of Physics: Condensed Matter **4** (1992), no. 32, 6651–6690.

[Bur99] ———, *Fundamentals of envelope function theory for electronic states and photonic modes in nanostructures*, Journal of Physics: Condensed Matter **11** (1999), no. 9, R53–R83.

[Bur02] V. A. Burdov, *Electron and Hole Spectra of Silicon Quantum Dots*, Journal of Experimental and Theoretical Physics **94** (2002), no. 2, 411–418.

[Büt86] M. Büttiker, *Four-Terminal Phase-Coherent Conductance*, Physical Review Letters **57** (1986), no. 14, 1761–1764.

[Büt88] ———, *Symmetry of electrical conduction*, IBM Journal of Research and Development **32** (1988), no. 3, 317–334.

[BUV05] M. Bayer, C. Uhl, and P. Vogl, *Theoretical study of electrolyte gate AlGaN/GaN field effect transistors*, Journal of Applied Physics **97** (2005), no. 3, 033703.

[BZA+07] S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, and P. Vogl, *nextnano: General Purpose 3-D Simulations*, IEEE Transactions on Electron Devices **54** (2007), no. 9, 2137–2142.

[Car03] X. Cartoixà, *Theoretical Methods for Spintronics in Semiconductors with Applications*, Ph.D. thesis, California Institute of Technology, Pasadena, California, USA, 2003.

[CC92] C. Y.-P. Chao and S. L. Chuang, *Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells*, Physical Review B **46** (1992), no. 7, 4110–4122.

[CC96] S. L. Chuang and C. S. Chang, $k \cdot p$ *method for strained wurtzite semiconductors*, Physical Review B **54** (1996), no. 4, 2491–2504.

[CP66] M. Cardona and F. H. Pollak, *Energy-Band Structure of Germanium and Silicon: The* $k \cdot p$ *Method*, Physical Review **142** (1966), no. 2, 530–543.

[CST] C. Schindler, private communication.

[CWPL01] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, *Nanowire Nanosensors for Highly Sensitive and Selective Detection of Biological and Chemical Species*, Science **293** (2001), no. 5533, 1289–1292.

[Dat95] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, Cambridge, 1995.

[Dat05] ———, *Quantum Transport: Atom to Transistor*, Cambridge University Press, Cambridge, 2005.

[DCVP94] A. Di Carlo, P. Vogl, and W. Pötz, *Theory of Zener tunneling and Wannier-Stark states in semiconductors*, Physical Review B **50** (1994), no. 12, 8358–8377.

[DDG+00] G. Dehlinger, L. Diehl, U. Gennser, H. Sigg, J. Faist, K. Ennslin, D. Grützmacher, and E. Müller, *Intersubband Electroluminescence from Silicon-Based Quantum Cascade Structures*, Science **290** (2000), no. 5500, 2277–2279.

[DHL+10] M. Dankerl, M. V. Hauf, A. Lippert, L. H. Hess, S. Birner, I. D. Sharp, A. Mahmood, P. Mallet, J.-Y. Veuillen, M. Stutzmann, and J. A. Garrido, *Graphene Solution-Gated Field-Effect Transistor Array For Sensing Applications*, Advanced Functional Materials **20** (2010), no. 18, 3117–3124.

[DKK55] G. Dresselhaus, A. F. Kip, and C. Kittel, *Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals*, Physical Review B **98** (1955), no. 2, 368–384.

[DLB+11] M. Dankerl, A. Lippert, S. Birner, E. U. Stützel, M. Stutzmann, and J. A. Garrido, *Hydrophobic interaction and charge accumulation at the diamond/electrolyte interface*, Physical Review Letters **106** (2011), no. 19, 196103.

[DLP+97] M. N. Darwish, J. L. Lentz, M. R. Pinto, P. M. Zeitzoff, T. J. Krutsick, and H. H. Vuong, *An Improved Electron and Hole Mobility Model for General Purpose Device Simulation*, IEEE Transactions on Electron Devices **44** (1997), no. 9, 1529–1538.

[DT94] D. P. Davé and H. F. Taylor, *Thomas-Kuhn sum rule for quantum mechanical systems with a spatially varying effective mass*, Physics Letters A **184** (1994), no. 3, 301–304.

[Eiß08] T. Eißfeller, *Linear Optical Response of Semiconductor Nanodevices*, Master's thesis, Walter Schottky Institut and Physics Department, Technische Universität München, Germany, 2008.

[EO56] H. Ehrenreich and A. W. Overhauser, *Scattering of Holes by Phonons in Germanium*, Physical Review **104** (1956), no. 2, 331–342.

[EPL10] E. T. Edmonds, C. I. Pakes, and L. Ley, *Self-consistent solution of the Schrödinger–Poisson equations for hydrogen-terminated diamond*, Physical Review B **81** (2010), no. 8, 085314.

[ESC87] R. Eppenga, M. F. H. Schuurmans, and S. Colak, *New $k \cdot p$ theory for $GaAs/Ga_{1-x}Al_xAs$-type quantum wells*, Physical Review B **36** (1987), no. 3, 1554–1564.

[ET70] L. Esaki and R. Tsu, *Superlattice and Negative Differential Conductivity in Semiconductors*, IBM Journal of Research and Development **14** (1970), no. 1, 61–65.

[EV11] T. Eißfeller and P. Vogl, *Spurious-solution-free real-space multi-band envelope function approach, submitted to* Physical Review B (2011).

[FAK+06] S. F. Fischer, G. Apetrii, U. Kunze, D. Schuh, and G. Abstreiter, *Tunnel-coupled one-dimensional electron systems with large subband separations*, Physical Review B **74** (2006), no. 11, 115324.

[FB89] R. Ferreira and G. Bastard, *Evaluation of some scattering times for electrons in unbiased and biased single- and multiple-quantum-well structures*, Physical Review B **40** (1989), no. 2, 1074–1086.

[FB03] V. A. Fonoberov and A. A. Balandin, *Excitonic properties of strained wurtzite and zinc-blende $GaN/Al_xGa_{1-x}N$ quantum dots*, Journal of Applied Physics **94** (2003), no. 11, 7178–7186.

[FG97] D. K. Ferry and S. M. Goodnick, *Transport in Nanostructures*, Cambridge University Press, Cambridge, 1997.

[FKXJ07] T. Fang, A. Konar, H. Xing, and D. Jena, *Carrier statistics and quantum capacitance of graphene sheets and ribbons*, Applied Physics Letters **91** (2007), no. 9, 092109.

[For93] B. A. Foreman, *Effective-mass Hamiltonian and boundary conditions for the valence bands of semiconductor microstructures*, Physical Review B **48** (1993), no. 7, 4964–4967.

[For97] ———, *Elimination of spurious solutions from eight-band $k \cdot p$ theory*, Physical Review B **56** (1997), no. 20, R12748–R12751.

[GFD+07] D. Grützmacher, T. Fromherz, C. Dais, J. Stangl, E. Müller, Y. Ekinci, H. H. Solak, H. Sigg, R. T. Lechner, E. Wintersberger, S. Birner, V. Holý, and G. Bauer, *Three-Dimensional Si/Ge Quantum Dot Crystals*, Nano Letters **7** (2007), no. 10, 3150–3156.

[GKTK99] E. Gheeraert, S. Koizumi, T. Teraji, and H. Kanda, *Electronic States of Boron and Phosphorus in Diamond*, physica status solidi (a) **174** (1999), no. 1, 39–51.

[GNHS08] J. A. Garrido, S. Nowy, A. Härtl, and M. Stutzmann, *The Diamond/Aqueous Electrolyte Interface: an Impedance Investigation*, Langmuir **24** (2008), no. 8, 3897–3904.

[Gra99] H. T. Grahn, *Introduction to Semiconductor Physics*, World Scientific Publishing Co. Pte. Ltd., Singapore, 1999.

[GU98] M. Governale and C. Ungarelli, *Gauge-invariant grid discretization of the Schrödinger equation*, Physical Review B **58** (1998), no. 12, 7816–7821.

[GYFE95] C. H. Grein, P. M. Young, M. E. Flatté, and H. Ehrenreich, *Long wavelength InAs/InGaSb infrared detectors: Optimization of carrier lifetimes*, Journal of Applied Physics **78** (1995), no. 12, 7143–7152.

[Hac02] S. Hackenbuchner, *Elektronische Struktur von Halbleiter-Nanobauelementen im thermodynamischen Nichtgleichgewicht*, Selected Topics of Semiconductor Physics and Technology (G. Abstreiter, M.-C. Amann, M. Stutzmann, and P. Vogl, eds.), vol. 48, Verein zur Förderung des Walter Schottky Instituts der Technischen Universität München e.V., München, 2002.

[HCM+09] I. Heller, S. Chatoor, J. Männick, M. A. G. Zevenbergen, C. Dekker, and S. G. Lemay, *Comparing the weak and strong gate-coupling regimes for nanotube and graphene transistors*, Physica Status Solidi RRL **3** (2009), no. 6, 190–192.

[HF63] J. C. Hensel and G. Feher, *Cyclotron Resonance Experiments in Uniaxially Stressed Silicon: Valence Band Inverse Mass Parameters and Deformation Potentials*, Physical Review **129** (1963), no. 5, 1041–1062.

[HK07] C. Heitzinger and G. Klimeck, *Computational aspects of the three-dimensional feature-scale simulation of silicon-nanowire field-effect sensors for DNA detection*, Journal of Computational Electronics **6** (2007), no. 1–3, 387–390.

[HS90] J. M. Hinckley and J. Singh, *Hole transport theory in pseudomorphic $Si_{1-x}Ge_x$ alloys grown on Si(001) substrates*, Physical Review B **41** (1990), no. 5, 2912–2926.

[HSNI83] S. Hiyamizu, J. Saito, K. Nanbu, and T. Ishikawa, *Improved Electron Mobility Higher than 10^6 cm^2/Vs in Selectively Doped GaAs/N-AlGaAs Heterostructures Grown by MBE*, Japanese Journal of Applied Physics **22** (1983), no. 10, L609–L611.

[HW78] T. W. Healy and L. R. White, *Ionizable surface group models of aqueous interfaces*, Advances in Colloid and Interface Science **9** (1978), no. 4, 303–345.

[JaM] J. Majewski, private communication.

[JG06] C.-W. Jiang and M. A. Green, *Silicon quantum dot superlattices: Modeling of energy bands, densities of states, and mobilities for silicon tandem solar cell applications*, Journal of Applied Physics **99** (2006), no. 11, 114902.

[JSBB98] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, *Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters*, Physical Review B **57** (1998), no. 11, 6493–6507.

[Kan69] E. O. Kane, Tunneling Phenomena in Solids (E. Burstein and S. Lundqvist, eds.), Plenum, New York, 1969.

[KAT01] L. P. Kouwenhoven, D. G. Austing, and S. Tarucha, *Few-electron quantum dots*, Reports on Progress in Physics **64** (2001), no. 6, 701–736.

[KDL09] R. Kim, S. Datta, and M. S. Lundstrom, *Influence of dimensionality on thermoelectric device performance*, Journal of Applied Physics **105** (2009), no. 3, 034506.

[KTV05] T. Kubis, A. Trellakis, and P. Vogl, *Self-consistent quantum transport theory of carrier capture in heterostructures*, Proceedings of the 14$^{\text{th}}$ International Conference on Nonequilibrium Carrier Dynamics in Semiconductors (HCIS 14, Chicago, USA), ed. by M. Saraniti and U. Ravaioli, Springer Proceedings in Physics **110** (2005), 369–372.

[Kub09] T. C. Kubis, *Quantum transport in semiconductor nanostructures*, Selected Topics of Semiconductor Physics and Technology (G. Abstreiter, M.-C. Amann, M. Stutzmann, and P. Vogl, eds.), vol. 114, Verein zur Förderung des Walter Schottky Instituts der Technischen Universität München e.V., München, 2009.

[KYV$^+$09] T. Kubis, C. Yeh, P. Vogl, A. Benz, G. Fasching, and C. Deutsch, *Theory of non-equilibrium quantum transport and energy dissipation in terahertz quantum cascade lasers*, Physical Review B **79** (2009), no. 19, 195323.

[Lan88] R. Landauer, *Spatial variation of currents and fields due to localized scatterers in metallic conduction*, IBM Journal of Research and Development **32** (1988), no. 3, 306–316.

[Lan92] ———, *Conductance from transmission: common sense points*, Physica Scripta **T42** (1992), 110–114.

[LAP] LAPACK – Linear Algebra PACKage, The LAPACK source code is available from `http://www.netlib.org/lapack`. Optimized libraries containing the LAPACK routines are included in several compiler suites.

[Law71] P. Lawaetz, *Valence-Band Parameters in Cubic Semiconductors*, Physical Review B **4** (1971), no. 10, 3460–3467.

[LB01] O. L. Lazarenkova and A. A. Balandin, *Miniband formation in a quantum dot crystal*, Journal of Applied Physics **89** (2001), no. 10, 5509–5515.

[LB02] ———, *Electron and phonon energy spectra in a three-dimensional regimented quantum dot superlattice*, Physical Review B **66** (2002), no. 24, 245319.

[LFC96] J. Los, A. Fasolino, and A. Catellani, *Generalization of the $k \cdot p$ approach for strained layered semiconductor structures grown on high-index-planes*, Physical Review B **53** (1996), no. 8, 4630–4648.

[Lip10] A. Lippert, *Electronic transport at the hydrogen-terminated diamond/ electrolyte interface*, Master's thesis, Walter Schottky Institut and Physics Department, Technische Universität München, Germany, 2010.

[LK55] J. M. Luttinger and W. Kohn, *Motion of Electrons and Holes in Perturbed Periodic Fields*, Physical Review **97** (1955), no. 2, 869–883.

[LK90] C. S. Lent and D. J. Kirkner, *The quantum transmitting boundary method*, Journal of Applied Physics **67** (1990), no. 10, 6353–6359.

[LKBJ97] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, *Single and multiband modeling of quantum electron transport through layered semiconductor devices*, Journal of Applied Physics **81** (1997), no. 12, 7845–7869.

[LKF04] S. E. Laux, A. Kumar, and M. V. Fischetti, *Analysis of quantum ballistic electron transport in ultrasmall silicon devices including space-charge and geometric effects*, Journal of Applied Physics **95** (2004), no. 10, 5545–5582.

[LNH+06] S. Q. Lud, M. G. Nikolaides, I. Haase, M. Fischer, and A. R. Bausch, *Field Effect of Screened Charges: Electrical Detection of Peptides and Proteins by a Thin-Film Resistor*, ChemPhysChem **7** (2006), no. 2, 379–384.

[Loe95] J. P. Loehr, *Parameter consistency in multienergetic $k \cdot p$ models*, Physical Review B **52** (1995), no. 4, 2374–2380.

[LSY98] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM Publications, Philadelphia, 1998.

[Lub06] S. M. Luber, *III-V semiconductor structures for biosensor and molecular electronics applications*, Selected Topics of Semiconductor Physics and Technology (G. Abstreiter, M.-C. Amann, M. Stutzmann, and P. Vogl, eds.), vol. 83, Verein zur Förderung des Walter Schottky Instituts der Technischen Universität München e.V., München, 2006.

[Lur88] S. Luryi, *Quantum capacitance devices*, Applied Physics Letters **52** (1988), no. 6, 501–503.

[Lut56] J. M. Luttinger, *Quantum Theory of Cyclotron Resonance in Semiconductors: General Theory*, Physical Review **102** (1956), no. 4, 1030–1039.

[LYVW09] L. C. Lew Yan Voon and M. Willatzen, *The $k \cdot p$ method – Electronic Properties of Semiconductors*, Springer, Berlin, 2009.

[MS96] K. Maschhoff and D. Sorensen, *P_ARPACK: An efficient portable large scale eigenvalue package for distributed memory parallel architectures*, Applied Parallel Computing Industrial Computation and Optimization (Jerzy Wasniewski, Jack Dongarra, Kaj Madsen, and Dorte Olesen, eds.), Lecture Notes in Computer Science, vol. 1184, Springer Berlin / Heidelberg, 1996, pp. 478–486.

[MSF07] L. G. Mourokh, A. Y. Smirnov, and S. F. Fischer, *Vertically coupled quantum wires in a longitudinal magnetic field*, Applied Physics Letters **90** (2007), no. 13, 132108.

[MSV03] D. Mamaluy, M. Sabathil, and P. Vogl, *Efficient method for the calculation of ballistic quantum transport*, Journal of Applied Physics **93** (2003), no. 8, 4628–4633.

[MU99] F. Mireles and S. E. Ulloa, *Ordered Hamiltonian and matching conditions for heterojunctions with wurtzite symmetry: $GaN/Al_xGa_{1-x}N$ quantum wells*, Physical Review B **60** (1999), no. 19, 13659–13667.

[MVS+05] D. Mamaluy, D. Vasileska, M. Sabathil, T. Zibold, and P. Vogl, *Contact block reduction method for ballistic transport and carrier densities of open nanostructures*, Physical Review B **71** (2005), no. 24, 245321.

[Net] R. R. Netz and N. Schwierz, private communication.

[NGM+04] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, *Electric Field Effect in Atomically Thin Carbon Films*, Science **306** (2004), no. 5696, 666–669.

[NPJJ+05] E. G. Novik, A. Pfeuffer-Jeschke, T. Jungwirth, V. Latussek, C. R. Becker, G. Landwehr, H. Buhmann, and L. W. Molenkamp, *Band structure of semimagnetic $Hg_{1-y}Mn_yTe$ quantum wells*, Physical Review B **72** (2005), no. 3, 035321.

[NRB04] M. G. Nikolaides, S. Rauschenbach, and A. R. Bausch, *Characterization of a silicon-on-insulator based thin film resistor in electrolyte solutions for sensor applications*, Journal of Applied Physics **95** (2004), no. 7, 3811–3815.

[NRL+03] M. G. Nikolaides, S. Rauschenbach, S. Luber, K. Buchholz, M. Tornow, G. Abstreiter, and A. R. Bausch, *Silicon-on-Insulator Based Thin-Film Resistor for Chemical and Biological Sensor Applications*, ChemPhysChem **4** (2003), no. 10, 1104–1106.

[OZV98] R. Oberhuber, G. Zandler, and P. Vogl, *Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFETs*, Physical Review B **58** (1998), no. 15, 9941–9948.

[PB66] C. R. Pidgeon and R. N. Brown, *Interband Magneto-Absorption and Faraday Rotation in InSb*, Physical Review **146** (1966), no. 2, 575–583.

[Ped06] D. Pedone, *Functionalized Silicon Nanowires for Labelfree Biosensing Applications*, Master's thesis, Walter Schottky Institut and Physics Department, Technische Universität München, Germany, 2006.

[PGBD+11] S. Prabhu-Gaunkar, S. Birner, S. Dasgupta, C. Knaak, and M. Grayson, *Valley degeneracy in biaxially strained aluminum arsenide quantum wells*, Physical Review B **84** (2011), no. 12, 125319 (15pp).

[PZ96] P. Pfeffer and W. Zawadzki, *Five-level $k \cdot p$ model for the conduction and valence bands of GaAs and InP*, Physical Review B **53** (1996), no. 19, 12813–12828.

[RAF04] S. Richard, F. Aniel, and G. Fishman, *Energy-band structure of Ge, Si, and GaAs: A thirty-band $k \cdot p$ method*, Physical Review B **70** (2004), no. 23, 235204.

[Rau61] C. J. Rauch, *Millimeter Cyclotron Resonance Experiments in Diamond*, Physical Review Letters **7** (1961), no. 3, 83–84.

[Rau62] ———, *Millimetre cyclotron resonance in diamond*, Proceedings of the International Conference on Semiconductor Physics, Exeter (The Institute of Physics and the Physical Society, London) (1962), 276–280.

[RK08] H. Ryu and G. Klimeck, *Contact Block Reduction Method for Ballistic Quantum Transport with Semi-empirical $sp^3d^5s^*$ Tight Binding Band Models*, 9th International Conference on Solid-State and Integrated-Circuit Technology (ICSICT 2008) (2008), 349–352.

[RMTO02] S. Reich, J. Maultzsch, C. Thomsen, and P. Ordejón, *Tight-binding description of graphene*, Physical Review B **66** (2002), no. 3, 035412.

[RWZ83] L. Reggiani, D. Waechter, and S. Zukotynski, *Hall-coefficient factor and inverse valence-band parameters of holes in natural diamond*, Physical Review B **128** (1983), no. 6, 3550–3555.

[Sab04] M. Sabathil, *Opto-electronic and quantum transport properties of semiconductor nanostructures*, Selected Topics of Semiconductor Physics and Technology (G. Abstreiter, M.-C. Amann, M. Stutzmann, and P. Vogl, eds.), vol. 67, Verein zur Förderung des Walter Schottky Instituts der Technischen Universität München e.V., München, 2004.

[SBMV03] M. Sabathil, S. Birner, D. Mamaluy, and P. Vogl, *Efficient Computational Method for Ballistic Currents and Application to Single Quantum Dots*, Journal of Computational Electronics **2** (2003), no. 2–4, 269–273.

[SC83] J. N. Schulman and Y. C. Chang, *Reduced Hamiltonian method for solving the tight-binding model of interfaces*, Physical Review B **27** (1983), no. 4, 2346–2354.

[Sca02] G. Scarpa, *Design and Fabrication of Quantum Cascade Lasers*, Selected Topics of Semiconductor Physics and Technology (G. Abstreiter, M.-C. Amann, M. Stutzmann, and P. Vogl, eds.), vol. 50, Verein zur Förderung des Walter Schottky Instituts der Technischen Universität München e.V., München, 2002.

[SCF94] C. Sirtori, F. Capasso, and J. Faist, *Nonparabolicity and a sum rule associated with bound-to-bound and bound-to-continuum intersubband transitions in quantum wells*, Physical Review B **50** (1994), no. 12, 8663–8674.

[Sch38] W. Schottky, *Halbleitertheorie der Sperrschicht*, Naturwissenschaften **26** (1938), no. 52, 843.

[SDT94] L. Schmitt, C. Dietrich, and R. Tampe, *Synthesis and Characterization of Chelator-Lipids for Reversible Immobilization of Engineered Proteins at Self-Assembled Lipid Interfaces*, Journal of the American Chemical Society **116** (1994), no. 19, 8485–8491.

[Sem04] Semiconductor Technology Research, Inc., *SiLENSe Version 1.0, Physics Summary*, `www.semitech.us`, Richmond, USA, 2004.

[SGB99] O. Stier, M. Grundmann, and D. Bimberg, *Electronic and optical properties of strained quantum dots modeled by 8-band $k \cdot p$ theory*, Physical Review B **59** (1999), no. 8, 5688–5701.

[SHN10] N. Schwierz, D. Horinek, and R. R. Netz, *Reversed Anionic Hofmeister Series: The Interplay of Surface Charge and Surface Polarity*, Langmuir **26** (2010), no. 10, 7370–7379.

[SK01] R. Saito and H. Kataura, *Optical Properties and Raman Spectroscopy of Carbon Nanotubes*, Carbon Nanotubes – Synthesis, Structure, Properties, and Applications (M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, eds.), Topics in Applied Physics, vol. 80, Springer New York, 2001, pp. 216–250.

[Smr90] L. Smrčka, *R-matrix and the coherent transport in mesoscopic systems*, Superlattices and Microstructures **8** (1990), no. 2, 221–224.

[SMV04] M. Sabathil, D. Mamaluy, and P. Vogl, *Prediction of a realistic quantum logic gate using the contact block reduction method*, Semiconductor Science and Technology **19** (2004), no. 4, S137–S138.

[SSRM06] Y. Shao, S. A. Solin, and L. R. Ram-Mohan, *Carrier Mobilities in delta-doped Heterostructures*, arXiv:cond-mat/0602140v1 (2006), 1–9.

[SSRMY07] Y. Shao, S. A. Solin, L. R. Ram-Mohan, and K.-H. Yoo, *Optimizing the physical contribution to the sensitivity and signal to noise ratio of extraordinary magnetoresistance quantum well structures*, Journal of Applied Physics **101** (2007), no. 12, 123704.

[Sze81] S. M. Sze, *Physics of Semiconductor Devices*, vol. 2nd ed., John Wiley & Sons, New York, 1981.

[TES] T. Eißfeller, private communication.

[TGPR97] A. Trellakis, A. T. Galick, A. Pacelli, and U. Ravaioli, *Iteration scheme for the solution of the two-dimensional Schrödinger–Poisson equations in quantum structures*, Journal of Applied Physics **81** (1997), no. 12, 7880–7884.

[TSCH90] I.-H. Tan, G. L. Snider, L. D. Chang, and E. L. Hu, *A self-consistent solution of Schrödinger–Poisson equations using a nonuniform mesh*, Journal of Applied Physics **68** (1990), no. 8, 4071–4076.

[TZA+06] A. Trellakis, T. Zibold, T. Andlauer, S. Birner, R. K. Smith, R. Morschl, and P. Vogl, *The 3D nanometer device project nextnano: Concepts, methods, results*, Journal of Computational Electronics **5** (2006), no. 4, 285–289.

[Uhl04] C. Uhl, *Charge Transport at Functionalized Semiconductor Surfaces*, Master's thesis, Walter Schottky Institut and Physics Department, Technische Universität München, Germany, 2004.

[VdW89] C. G. Van de Walle, *Band lineups and deformation potentials in the model-solid theory*, Physical Review B **39** (1989), no. 3, 1871–1883.

[VK10] P. Vogl and T. Kubis, *The non-equilibrium Green's function method: an introduction*, Journal of Computational Electronics **9** (2010), no. 3–4, 237–242.

[VM03] I. Vurgaftman and J. R. Meyer, *Band parameters for nitrogen-containing semiconductors*, Journal of Applied Physics **94** (2003), no. 6, 3675–3696.

[VMK+08] D. Vasileska, D. Mamaluy, H. R. Khan, K. Raleva, and S. M. Goodnick, *Semiconductor Device Modeling*, Journal of Computational and Theoretical Nanoscience **5** (2008), no. 6, 999–1030.

[VMRM01] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, *Band parameters for III-V compound semiconductors and their alloys*, Journal of Applied Physics **89** (2001), no. 11, 5815–5875.

[VSW07] R. G. Veprek, S. Steiger, and B. Witzigmann, *Ellipticity and the spurious solution problem of $k \cdot p$ envelope functions*, Physical Review B **76** (2007), no. 16, 165320.

[Wal74] P. R. Wallace, *The Band Theory of Graphite*, Physical Review **71** (1974), no. 9, 622–634.

[WCC94] M. Willatzen, M. Cardona, and N. E. Christensen, *Linear muffin-tin-orbital and $k \cdot p$ calculations of effective masses and band structure of semiconducting diamond*, Physical Review B **50** (1994), no. 24, 18054–18059.

[WLLYVM06] M. Willatzen, B. Lassen, L. C. Lew Yan Voon, and R. V. N. Melnik, *Dynamic coupling of piezoelectric effects, spontaneous polarization, and strain in lattice-mismatched semiconductor quantum-well heterostructures*, Journal of Applied Physics **100** (2006), no. 2, 024302.

[WRLG84] W. Walukiewicz, H. E. Ruda, J. Lagowski, and H. C. Gatos, *Electron mobility in modulation-doped heterostructures*, Physical Review B **30** (1984), no. 8, 4571–4582.

[www] The next**nano**^{3} and next**nano**++ software can be obtained from `http://www.nextnano.de`.

[WZ96] D. M. Wood and A. Zunger, *Successes and failures of the $k \cdot p$ method: A direct assessment for GaAs/AlAs quantum structures*, Physical Review B **53** (1996), no. 12, 7949–7963.

[XZP11] H. Xu, Z. Zhang, and L.-M. Penga, *Measurements and microscopic model of quantum capacitance in graphene*, Applied Physics Letters **98** (2011), no. 13, 133122.

[YC99] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors: Physics and Materials Properties*, second ed., Springer, Berlin, 1999.

[Zib07] T. Zibold, *Semiconductor based quantum information devices: Theory and simulations*, Selected Topics of Semiconductor Physics and Technology (G. Abstreiter, M.-C. Amann, M. Stutzmann, and P. Vogl, eds.), vol. 87, Verein zur Förderung des Walter Schottky Instituts der Technischen Universität München e.V., München, 2007.

[ZVB07] T. Zibold, P. Vogl, and A. Bertoni, *Theory of semiconductor quantum-wire-based single- and two-qubit gates*, Physical Review B **76** (2007), no. 19, 195301.

[ZYC01] A. Zakharova, S. T. Yen, and K. A. Chao, *Hybridization of electron, light-hole, and heavy-hole states in InAs/GaSb quantum wells*, Physical Review B **64** (2001), no. 23, 235332.