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

Optoelectronic and spin-related properties of semiconductor nanostructures in magnetic fields

Till Andlauer

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. Dr. Wilhelm Zwerger

Die Dissertation wurde am 20.01.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 11.03.2009 angenommen.

Contents

Abstract

In this work, the electronic structure of realistic semiconductor nanostructures subject to external magnetic fields is calculated. To this end, a gauge-invariant nonperturbative discretization scheme is developed for the multiband $\mathbf{k}\cdot\mathbf{p}$ envelope function theory including strain as well as relativistic effects. The method avoids the gauge dependent spectrum of a straightforwardly discretized minimal coupling Hamiltonian by transferring Wilson's formulation of lattice gauge theories to the field of semiconductor physics. With this procedure, electron and hole spin g factors of InAs/InP nanowire-based and InAs/GaAs self-assembled quantum dots are calculated. Excellent agreement with recent experimental data verifies the accuracy of the approach. For a concrete application proposal, the analysis is extended to vertically coupled InAs/GaAs quantum dot pairs in external electric fields. For magnetic fields lying in the growth plane, a giant electrically tunable anisotropy of hole g factors is predicted that is introduced by piezoelectric charges. This effect allows bias controlled g factor switching and single-spin manipulations in a static magnetic field. In a regime where the molecular wave functions form bonding and antibonding orbitals and for vertical magnetic fields, the calculations reproduce experimentally observed resonant enhancements of exciton g factors without any fitting parameters.

A second important topic of this work are optoelectronic properties of broken-gap heterostructures. Standard effective mass theory fails to yield the correct occupation of electronic states in heterostructures with a type-II broken-gap band alignment, because the strong hybridization of conduction and valence bands prevents an a-priori classification into electron and hole states. Therefore, a novel charge self-consistent electronic structure scheme is developed that remains in the electron framework throughout. Applying this procedure, optical transition energies are calculated for a series of intrinsic InAs/GaSb superlattices with different layer widths.

Finally, the electronic structure of shallow impurities is studied in close collaboration with an experimental group. In this context, the energy shifts of acceptor Zeeman levels in bulk silicon are calculated via isotope induced local fluctuations of band gap energies. In addition, the contributions of quantum confinement and dielectric screening to the localization of donor wave functions in silicon nanocrystals are investigated.

Zusammenfassung

In dieser Arbeit wird die elektronische Struktur von realistischen Halbleiter Nanostrukturen in externen magnetischen Feldern berechnet. Zu diesem Zweck wird ein eichinvariantes, nicht-störungstheoretisches Diskretisierungsverfahren für die Mehrband $\mathbf{k}\cdot\mathbf{p}$ Einhüllenden-Funktions-Theorie entwickelt. Die Methode umgeht die Eichabhängigkeit des Spektrums eines direkt diskretisierten Hamilton Operators auf Basis der minimalen Kopplung, indem Wilsons Formulierung der Gittereichtheorie auf das Gebiet der Halbleiterphysik übertragen wird. Mit dieser Methode werden Elektronen und Löcher Spin g Faktoren von InAs/InP Nanodraht basierten und InAs/GaSb selbstorganisierten Quantenpunkten berechnet. Die ausgezeichnete Übereinstimmung mit neuesten experimentellen Daten bestätigt die Genauigkeit des Verfahrens. Zur Erarbeitung eines konkreten Anwendungsvorschlags wird die Untersuchung auf senkrecht gekoppelte Paare von InAs/GaAs Quantenpunkten in externen elektrischen Feldern ausgedehnt. Für den Fall, dass das magnetische Feld in der Wachstumsebene liegt, wird eine starke, elektrisch einstellbare Richtungsabhängigkeit der Löcher g Faktoren vorhergesagt, die durch piezoelektrische Ladungen ausgelöst wird. Dieser Effekt erlaubt es g Faktoren spannungsgesteuert zu schalten und ermöglicht Einzel-Spin Manipulationen im statischen Magnetfeld. In einem Bereich, in dem die molekularen Wellenfunktionen bindende und antibindende Orbitale bilden und für senkrechte Magnetfeldrichtung, reproduzieren die Berechnungen experimentell beobachtete, resonante Verstärkungen von Exziton g Faktoren, ohne Verwendung jeglicher Anpassungsparameter.

Ein zweiter Schwerpunkt dieser Arbeit liegt auf den optoelektronischen Eigenschaften von Heterostrukturen mit unterbrochener Bandlücke. Die Standard Effektive-Masse-Theorie scheitert daran, die korrekte Besetzung der elektronischen Zustände für Heterostrukturen mit Typ-II Bandkanten Anordnung zu bestimmen. Dies ist durch die starke Hybridisierung von Leitungs- und Valenzbändern zu erklären, die eine a priori Klassifizierung in Elektron- und Lochzustände verhindert. Daher wird ein neuartiges, ladungsselbstkonsistentes Elektronische-Struktur-Verfahren entwickelt, das durchgängig im Elektronenbild bleibt. Mit dieser Methode werden optische Übergangsenergien für eine Reihe von eigenleitenden InAs/GaSb Übergittern mit verschiedenen Schichtdicken berechnet.

Abschließend wird in enger Zusammenarbeit mit einer experimentellen Gruppe die elektronische Struktur von flachen Störstellen untersucht. In diesem Zusammenhang, werden die Energieverschiebungen von Akzeptor Zeeman-Niveaus in natürlichem Silizium über Isotopen verursachte lokale Änderungen der Bandlücke berechnet. Desweiteren werden die Einflüsse von quantenmechanischer Lokalisierung und dielektrischer Abschirmung auf die Donator Wellenfunktionen in Silizium Nanokristallen überprüft.

Introduction

The advances in semiconductor technology have made it possible, to fabricate devices that are structured on a nanometer scale. This opens up promising possibilities for the development of novel electronic and optical devices that rely on quantum mechanical effects and consume much less power or allow much faster computations. Rather than using electric charges, many next generation device concepts rely on the spin degree of freedom for the logical units in quantum computation [1] or for carrying information in spintronics [2]. Free carriers in low dimensional semiconductor nanostructures are promising candidates that provide a natural two-level spin-1/2 system and allow good scalability. In many concepts developed so far, the spins are being manipulated by external magnetic fields. Obviously, progress in this field requires a detailed understanding of the mechanisms that allow one to control spin-related electronic structure properties such as gyromagnetic factors (also called g factors). In addition, there is a growing demand for environmental sensing and fiber-optic communications, both relying on infrared lasers and infrared detectors. There is a whole family of optoelectronic device concepts for this spectral regime that is based on the unique properties of semiconductor tunnel junctions with the lowest conduction band in one material layer energetically lying below the top valence band in an adjacent layer [3]. However, the electronic structure properties of the respective material system differ fundamentally from a conventional semiconductor and are not yet fully understood. Next generation electronic and optical devices in these fields are currently being developed world wide. However, experimental realization of existing proposals for such devices is still a complicated task. Simple analytical models may help in understanding the basic physical relations, but they are not sufficient to support the concrete realization and optimization of nanodevices. Numerical tools that allow a predictive quantitative analysis of realistic three-dimensional structures are therefore strongly desirable.

In this work, we have developed a simulation tool that provides a global insight into a wide range of electronic, optical, and transport characteristics of mesoscopic structures with virtually any geometry and combination of semiconducting materials. It focuses on quantum mechanical properties such as the global electronic structure, optical properties, and the effects of electric and magnetic fields. Only recently, it has been recognized that an accurate incorporation of the magnetic field into the Schrödinger equation requires special care to ensure gauge-invariant results.

Therefore, a gauge-invariant discretization scheme is developed for the multiband $\mathbf{k}\cdot\mathbf{p}$ envelope function theory. This procedure is used for a quantitative investigation of electron and hole gyromagnetic tensors in realistic three-dimensional semiconductor nanostructures including the detailed geometry and material composition. The goal of this study is to propose concrete viable nanostructures that allow to efficiently control single spins. In this point, the present work differs fundamentally from the large number of approaches, which provide only basic device concepts using simple physical models or idealized geometries. A second focus of this work lies on the investigation of optoelectronic properties of heterostructures that do not have a global band gap. To this end, a novel electronic structure scheme is developed that maintains the efficiency of a continuum approach yet does not depend on a separation into negatively charged electron and positively charged hole states, which would fail due to the strong hybridization of conduction and valence bands. With this method, the technologically important InAs/GaSb material system is studied.

This thesis is organized as follows. We start with a description of our semiconductor nanodevice simulation package next**nano**++ [4]. Its development has been an important part of the present work and marks the basis for all more advanced calculations. In chapter 1, we introduce the underlying physical concepts needed for the calculation of semiconductor nanostructures. Here, the basic physical equations for the calculation of the electronic structure, the external potentials, the elastic strain, and the electric currents are described. For the numerical realization of such computationally demanding calculations, the development of novel numerical concepts has been necessary. In chapter 2, we describe these highly efficient computational methods that allow one to calculate the properties of three-dimensional semiconductor nanostructures even on a standard off-the-shelf PC. In this chapter, we also sketch the design of next**nano**++ applying modern object oriented programming techniques on a modular code setup. After these fundamental introductory parts, we turn to the key point of this thesis, namely the investigation of semiconductor nanostructures in external magnetic fields. In order to perform these calculations, a novel method is developed in chapter 3 for the multiband $\mathbf{k}\cdot\mathbf{p}$ envelope function theory that includes the coupling to the magnetic field in a manifestly gauge-invariant manner. The procedure is applied in chapter 4 to calculate magnetic-field related electronic structure properties of a large variety of quantum dots. In chapter 5, the investigation is extended to coupled quantum dots that provide more room for electric tuning of the Larmor precession, which is a key requirement for fast quantum gate operations. Indeed, we find promising electrically controllable g tensors that allow coherent single-spin manipulations in a static magnetic field. Next, we turn to the optoelectronic properties of nanostructures with a type-II broken-gap band alignment. In chapter 6, we present a novel charge self-consistent multiband envelope function approach for broken-gap heterostructures and calculate optical transition energies for a series of the most typical superlattices. Chapter 7 forms the third major application part of this thesis, where we study electronic structure properties of shallow impurities in close collaboration with the experimental group of Prof. Martin Brandt. This analysis is important to improve the understanding of spin states for silicon based quantum computers as well as electronic doping at the nanoscale. Finally, the thesis is summarized in chapter 8 and an outlook is given.

Chapter 1

Calculation of semiconductor nanostructures

1.1 Introduction

In this chapter, we address the theoretical concepts that are used in next**nano**++ [4] for realistic three-dimensional calculations of semiconductor nanostructures. Since the software aims at providing global insight into the basic physical properties of mesoscopic semiconductor structures, it requires the modeling of a large variety of physical aspects. Here, a focus lies on quantum mechanical properties that require the choice of an adequate electronic structure model. In realistic systems, there are many connections between different physical properties. Thus, it is essential to find suitable models for the different aspects that can be combined in a consistent way.

Common theoretical electronic structure principles can be grouped into continuum and atomistic approaches. The latter ones are more sensitive to the underlying crystal structure but also computationally much more demanding. Continuum approaches have the advantage of being scalable to larger devices without excessive increase of numerical effort. Concrete realizations of semiconductor nanostructures are usually embedded in micrometer scale semiconductor environments such as a substrate or electric contacts. Since there are possibly long range effects (e.g. Coulomb forces and strain fields), it is often necessary to actually calculate an extended mesoscopic system. In next**nano**++, we therefore completely rely on continuum models for all types of physical aspects. This has the advantage that the appropriate differential equations can be mapped on the same inhomogenous scalable grid where they can be combined in an unambiguous and consistent manner.

The software next**nano**++ is based on concepts previously introduced in the code next**nano**^{3} [5], but involves rigorous improvements both from a physical and a numerical point of view. The new code has been developed in close collaboration with Tobias Zibold and Alex Trellakis. Many features of next**nano**++ have therefore been described in Ref. [6]. In this chapter and the subsequent one, further details will be presented on parts where the author has been directly involved in the implementation. Concretely, the relativistic multiband **k·p** envelope function method is discussed that is employed to calculate the global quantum mechanical electronic structure. Magnetic fields can be incorporated in this Hamiltonian, but a detailed description of the procedure will be given separately in chapter 3. To take into account free-carrier charges, doping, and fixed charges, the Hamiltonian is augmented by the Hartee potential that is obtained from the Poisson equation. Exchange and correlation effects are included by means of the local spin density approximation. Electric fields can be applied via boundary conditions in the Poisson equation. Strain effects are incorporated by linear band-edge deformation potentials and piezoelectric charges. The strain field is calculated by minimizing the total elastic energy in a continuum elasticity model. For the calculation of carrier dynamics, two models are currently implemented that provide results for the limiting cases of highly diffusive or purely ballistic quantum-mechanical transport. In this work, we will only discuss the former one, namely a quantum drift-diffusion model that has been combined with the Schrödinger- and Poisson equations in a fully self-consistent manner. Readers interested in the ballistic transport model (named CBR method) are referred to Refs. [6, 7].

1.2 Band structure

1.2.1 Multiband k·p envelope function method

Semiconductor nanostructures exhibit a manifestly quantum mechanical behavior in terms of their electronic properties. Therefore, we have to solve the Schrödinger equation for the entire nanodevices to compute their global electronic structure. We rely on the multiband **k·p** method together with the envelope function approximation (EFA) [8–19]. The basic idea of this method is to patch up the bulk **k·p** Hamiltonian of each constituent material such that the global Hamiltonian remains Hermitian. While this method is inferior to microscopic electronic structure methods such as pseudopotential or empirical tight-binding schemes on an atomic scale, it is the method of choice for structures that extend over many tens or hundreds of nanometers and reflects accurately the extended electronic states in a mesoscopic device.

Basic k·p theory

In general, the electronic structure of a bulk semiconductor is determined by a many-particle Hamiltonian that includes all electrons and nuclei of the crystal. It is well known from standard textbooks on solid-state physics [10, 20] that by assuming the atomic cores to be stationary (adiabatic approximation) and by applying a mean field approximation for all multi-particle interactions, the problem can effectively be reduced to a single-particle Schrödinger equation of an electron in a perfectly periodic crystal

$$\hat{H}\psi_n(\mathbf{x}) = \left[\frac{\mathbf{p}^2}{2m_0} + V(\mathbf{x})\right]\psi_n(\mathbf{x}) = E_n\psi_n(\mathbf{x}). \tag{1.1}$$

Here, $V(\mathbf{x})$ is a mean field potential that includes all interactions and reflects the translation invariance of the crystal

$$V(\mathbf{x}+\mathbf{R}) = V(\mathbf{x}), \tag{1.2}$$

where $\mathbf{R}$ is a Bravais lattice vector. The eigenfunctions $\psi_n(\mathbf{x})$ of Eq. (1.1) obey Bloch's theorem, i.e., they are at the same time eigenfunctions $\psi_{n,\mathbf{k}}(\mathbf{x})$ of the translation operator $\hat{T}_{\mathbf{R}}$

$$\hat{T}_{\mathbf{R}}\psi_{n,\mathbf{k}}(\mathbf{x}) = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{R})\,\psi_{n,\mathbf{k}}(\mathbf{x}), \tag{1.3}$$

and can be written as Bloch functions

$$\psi_{n,\mathbf{k}}(\mathbf{x}) = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})\,u_{n,\mathbf{k}}(\mathbf{x}), \tag{1.4}$$

with the periodic Bloch factors $u_{n,\mathbf{k}}(\mathbf{x})$ and plane waves $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})$. Inserting these Bloch functions into Schrödinger's equation [Eq. (1.1)] leads to the relation

$$\hat{H}(\mathbf{k})\,u_{n,\mathbf{k}}(\mathbf{x}) = \left[\frac{(\mathbf{p}+\hbar\mathbf{k})^2}{2m_0} + V(\mathbf{x})\right]u_{n,\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\,u_{n,\mathbf{k}}(\mathbf{x}). \tag{1.5}$$

Now, the remaining problem is the calculation of the band structure $E_n(\mathbf{k})$, which is the dependence of the band eigenenergies E_n on the wave vector $\mathbf{k}$ in the first Brillouin zone. Numerous methods have been developed to perform this task. While a straightforward way would be to find a good approximation for the mean field potential $V(\mathbf{x})$ in order to solve Eq. (1.5), the $\mathbf{k}\cdot\mathbf{p}$ method follows a different approach. It utilizes the fact that carriers in semiconductors usually occupy only regions close to the minima and maxima of the conduction and valence bands and a precise knowledge of the remaining parts of the band structure is unnecessary. To this end, the $\mathbf{k}\cdot\mathbf{p}$ method expands the band structure around a certain extremum $\mathbf{k}_0$, where the energies $E_n(\mathbf{k}_0)$ and Bloch factors $u_{n,\mathbf{k}_0}(\mathbf{x})$ are assumed to be known.

So we split the Hamiltonian $\hat{H}(\mathbf{k})$ [Eq. (1.5)] into a constant extremum part $\hat{H}(\mathbf{k}_0)$ and the remaining $\mathbf{k}$-dependent parts as follows

$$\hat{H}(\mathbf{k}) = \hat{H}(\mathbf{k}_0) + \frac{\hbar(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{p}}{m_0} + \frac{\hbar^2(\mathbf{k}^2-\mathbf{k}_0^2)}{2m_0}. \tag{1.6}$$

The extremum part $\hat{H}(\mathbf{k}_0)$ has the solution

$$\hat{H}(\mathbf{k}_0)\,u_{n,\mathbf{k}_0}(\mathbf{x}) = E_n(\mathbf{k}_0)\,u_{n,\mathbf{k}_0}(\mathbf{x}), \tag{1.7}$$

with Bloch factors $u_{n,\mathbf{k}_0}(\mathbf{x})$ that form a complete and orthonormal basis. Thus we can expand the unknown Bloch factors $u_{n,\mathbf{k}}(\mathbf{x})$ for any value of $\mathbf{k}$ in terms of the known $u_{n,\mathbf{k}_0}(\mathbf{x})$,

$$u_{n,\mathbf{k}}(\mathbf{x}) = \sum_{\mu} c_{n,\mu}(\mathbf{k})\, u_{\mu,\mathbf{k}_0}(\mathbf{x}) \,. \tag{1.8}$$

Next, we insert Eq. (1.8) into the Schrödinger equation [Eq. (1.5)] with the Hamiltonian of Eq. (1.6). After multiplying both sides with $u_{\nu,\mathbf{k}_0}(\mathbf{x})$ and integrating over the unit cell, we obtain an eigenvalue equation for the expansion coefficients

$$\sum_{\mu} \hat{H}_{\nu\mu}(\mathbf{k})\, c_{n,\mu}(\mathbf{k}) = E_n(\mathbf{k})\, c_{n,\nu}(\mathbf{k}) \,, \tag{1.9}$$

with the Hamiltonian given by the matrix

$$\hat{H}_{\nu\mu}(\mathbf{k}) = \left[E_\mu(\mathbf{k}_0) + \frac{\hbar^2}{2m_0}\left(\mathbf{k}^2 - \mathbf{k}_0^2\right) \right] \delta_{\nu\mu} + \frac{\hbar}{m_0}(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{p}_{\nu\mu}. \tag{1.10}$$

Up to now, no approximations have been made and $E_n(\mathbf{k})$ is exact for any band n and wave vector $\mathbf{k}$. However, the Hamiltonian $\hat{H}_{\nu\mu}(\mathbf{k})$ is an infinite dimensional matrix that couples all energy bands via the momentum matrix elements

$$\mathbf{p}_{\nu\mu} = \int_{\Omega} d\mathbf{x}\, u^*_{\nu,\mathbf{k}_0}(\mathbf{x})\; \mathbf{p}\, u_{\mu,\mathbf{k}_0}(\mathbf{x}) \,. \tag{1.11}$$

Next, one utilizes the fact that for many purposes only a few bands and only wave vectors in the vicinity of the extrema of these bands are physically relevant. The number of bands is reduced by employing Löwdin's perturbation theory [21], which allows to decouple a finite set of bands in class A from the remaining bands in class B. Customarily, class A only contains conduction and valence bands that lie close to the Fermi energy, while all other bands are grouped in class B. The range of $\mathbf{k}$ values that give accurate results for $E_n(\mathbf{k})$ is restricted by treating the off-diagonal part of the Hamiltonian

$$\frac{\hbar}{m_0}(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{p}_{\nu\mu}, \tag{1.12}$$

as a perturbation. Note that for the decoupling transformation, the energy separation among class A bands is considered to be much smaller than the energy separation between any class A and class B band. Altogether, the infinite dimensional Hamiltonian matrix $\hat{H}_{\nu\mu}(\mathbf{k})$ [Eq. (1.10)] is transformed into a finite dimensional matrix $\hat{H}'_{\nu\mu}(\mathbf{k})$ with renormalized coupling constants,

$$\hat{H}'_{\nu\mu}(\mathbf{k}) = \hat{H}_{\nu\mu}(\mathbf{k}) + \sum_{\beta \in B} \frac{\hat{H}_{\nu\beta}(\mathbf{k})\, \hat{H}_{\beta\mu}(\mathbf{k})}{E_\nu(\mathbf{k}_0) - E_\beta(\mathbf{k}_0)}, \tag{1.13}$$

where ν, μ now run only over the N_A bands in class A. Typically, the resulting N_A-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian is fully determined by only a few independent band-edge energies and momentum matrix elements that can be found by considering the symmetry of the underlying crystal [11]. These elements are commonly used as empirical parameters that are fitted to match experimental properties of the bulk band structure [22]. In the following section, we will present in some detail those **k**·**p** models that we have implemented within next**nano**++. We restrict the discussion to the zincblende crystal structure, which is suitable to describe Si, Ge, and most III-V semiconductors. The corresponding models for the less common wurtzite crystals (basically nitrides) will be summarized in Appendix A.

Hamiltonian

Single-band models In the simplest model, which is called effective mass approximation (EMA), only a single band is considered in class A. Here, the Hamiltonian matrix of Eq. (1.13) is reduced to a scalar ($\nu, \mu = n$)

$$\begin{aligned} \hat{H}_n^{\mathrm{EMA}}(\mathbf{k}) &= E_n(\mathbf{k}_0) + \frac{\hbar^2}{2m_0}\left(\mathbf{k}^2 - \mathbf{k}_0^2\right) + \frac{\hbar^2}{m_0^2}\sum_{\beta\neq n}\frac{\left|(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{p}_{n\beta}\right|^2}{E_n(\mathbf{k}_0) - E_\beta(\mathbf{k}_0)} \\ &= E_n(\mathbf{k}_0) + \frac{\hbar^2}{2}(\mathbf{k}-\mathbf{k}_0)^T\frac{1}{\hat{m}^*}(\mathbf{k}-\mathbf{k}_0), \end{aligned} \tag{1.14}$$

with an effective mass tensor $\hat{m}^*$ that is a symmetric 3×3 matrix. For a zincblende conduction band, there can be local minima at several high symmetry points of the Brillouin zone, namely the zone center Γ-point ($\mathbf{k}_0 = 0$), the six X-points ($\mathbf{k}_0 = \{100\}$), and the eight L-points ($\mathbf{k}_0 = \{111\}/\sqrt{3}$). In silicon, there are minima lying along the six Δ-lines (connections between Γ and X), rather than directly at the X-points. Note that the individual X- and L-points all lie at the boundary of the first Brillouin zone. Thus, each of them is shared by two adjacent Brillouin zones and only half of them have to be considered in the charge density calculation (cf. Sec. 1.3.1). The effective mass tensor of the Γ-valley is isotropic and can be described by a scalar effective mass m^*. For the L-, X-, and Δ-valleys, the effective mass tensor is characterized by a longitudinal mass m_l^* and a transversal mass m_t^*,

$$\hat{m}^* = \left(m_l^* - m_t^*\right)\mathbf{e}_{\mathbf{k}_0}\mathbf{e}_{\mathbf{k}_0}^T + m_t^* 1^{3\times 3}. \tag{1.15}$$

Here, $\mathbf{e}_{\mathbf{k}_0}$ is a unit vector that points in the direction of the individual minimum $\mathbf{k}_0$. The anisotropic effective mass can be best described by an ellipsoid with the symmetry axis lying in the $\mathbf{e}_{\mathbf{k}_0}$-direction. In strained semiconductors, we add a strain-dependent Hamiltonian $\hat{H}_\varepsilon$ [Eq. (1.89)] to Eq. (1.14). It shifts the band-edge energies via deformation potentials and will be discussed in Sec. 1.4.2.

For the conduction-band minima of wide gap semiconductors, the EMA is often sufficiently accurate. For the valence-bands and for the Γ-conduction band of narrow gap materials, however, we usually rely on more sophisticated multiband models. This has been exemplified in Fig. 1.1 that shows a typical band structure of a narrow gap semiconductor. In the present work, it will often be referred to InAs which is a good example for such a material.

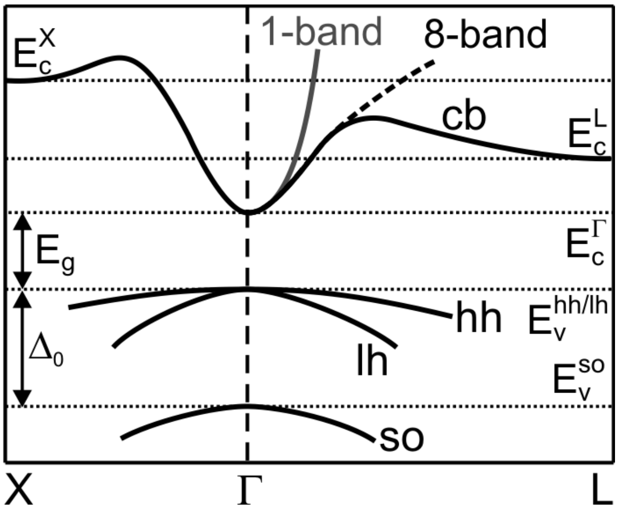


Figure 1.1: Schematic band structure of a narrow direct gap semiconductor. The conduction band has minima at the Γ-, X-, and L-points. The heavy- and light hole valence bands (hh/lh) are degenerate at the Γ-point, while the split-off band (so) is shifted in energy by the spin-orbit splitting Δ_0. Due to strong nonparabolicities, a single-band model (solid gray line) only poorly describes the conduction band. For such a material, an eight-band model (dashed black line) that includes the coupling to the valence bands, gives a significantly more accurate approximation.

Multiband models Since the highest valence band Bloch functions are p orbitals, they are threefold degenerate. In addition, they encounter a strong spin-orbit coupling that has to be taken into account by adding the relativistic correction term

$$\hat{H}_{\mathrm{so}} = \frac{\hbar}{4m^2c^2} \left(\boldsymbol{\nabla} V \times \mathbf{p} \right) \cdot \boldsymbol{\sigma}, \tag{1.16}$$

to the single particle Hamiltonian in Eq. (1.1). Here $\boldsymbol{\sigma}$ denotes a vector that is built up by the Pauli spin matrices σ_i. By this, the solutions of Schrödinger's equation become two-component spinors and the group of bands A in Eq. (1.13) needs to comprise at least six bands. In the basis of spin-resolved zone center valence-band Bloch functions $|q\sigma\rangle \in \{x_1, x_2, x_3\} \otimes \{\uparrow, \downarrow\}$, the six-band $\mathbf{k}{\cdot}\mathbf{p}$ Hamiltonian resulting from Eq. (1.13) can be written in the following form [8, 17, 19]

$$\hat{H}^{6\times 6}(\mathbf{k}) = \begin{pmatrix} \hat{H}^{3\times 3}(\mathbf{k}) & 0 \\ 0 & \hat{H}^{3\times 3}(\mathbf{k}) \end{pmatrix} + \hat{H}^{6\times 6}_{\mathrm{so}}, \tag{1.17}$$

$$\hat{H}^{3\times 3}(\mathbf{k}) = \left(E_v + \frac{\hbar^2}{2m_0}\sum_{i=1}^{3} k_i^2 \right) \otimes 1^{3\times 3}$$
$$+ \begin{pmatrix} k_1 L k_1 + \sum\limits_{i=2,3} k_i M k_i & k_1 N_+ k_2 + k_2 N_- k_1 & k_1 N_+ k_3 + k_3 N_- k_1 \\ k_1 N_- k_2 + k_2 N_+ k_1 & k_2 L k_2 + \sum\limits_{i=1,3} k_i M k_i & k_2 N_+ k_3 + k_3 N_- k_2 \\ k_1 N_- k_3 + k_3 N_+ k_1 & k_2 N_- k_3 + k_3 N_+ k_2 & k_3 L k_3 + \sum\limits_{i=1,2} k_i M k_i \end{pmatrix}, \tag{1.18}$$

with $E_v = E_v(\mathbf{k}_0 = 0)$. Instead of the original linearly independent Dresselhaus parameters F, G, H_1, H_2 [23], we have specified this Hamiltonian in terms of the more commonly used derived Dresselhaus parameters L, M, N_+, N_-, that are related to each other by

$$\begin{aligned} L = F + 2G,\ M = H_1 + H_2, \\ N_+ = F - G,\ N_- = H_1 - H_2. \end{aligned} \tag{1.19}$$

The latter two parameters have been introduced by Ref. [19]. Another commonly used and tabulated set of derived Luttinger parameters is $\gamma_1, \gamma_2, \gamma_3, \kappa$ [12, 24, 25]. In terms of these parameters, we can write

$$\begin{aligned} L = \frac{\hbar^2}{2m_0}\left(-\gamma_1 - 4\gamma_2 - 1\right),\ M = \frac{\hbar^2}{2m_0}\left(-\gamma_1 + 2\gamma_2 - 1\right), \\ N_+ = \frac{\hbar^2}{2m_0}\left(-3\gamma_3 - 3\kappa - 1\right),\ N_- = \frac{\hbar^2}{2m_0}\left(-3\gamma_3 + 3\kappa + 1\right). \end{aligned} \tag{1.20}$$

The spin-orbit Hamiltonian is given by

$$\hat{H}_{\mathrm{so}}^{6\times 6} = \frac{\Delta_0}{3}\begin{pmatrix} 0 & -\mathrm{i} & 0 & 0 & 0 & 1 \\ \mathrm{i} & 0 & 0 & 0 & 0 & -\mathrm{i} \\ 0 & 0 & 0 & -1 & \mathrm{i} & 0 \\ 0 & 0 & -1 & 0 & \mathrm{i} & 0 \\ 0 & 0 & -\mathrm{i} & -\mathrm{i} & 0 & 0 \\ 1 & \mathrm{i} & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{1.21}$$

where Δ_0 is the spin-orbit splitting. Here, we have neglected the small k-dependent contribution $(\hbar/4m^2c^2)(\boldsymbol{\nabla} V \times \mathbf{k})\cdot\boldsymbol{\sigma}$ that results from employing the $\mathbf{k}\cdot\mathbf{p}$ approximation on Eq. (1.16). The spin-orbit Hamiltonian $\hat{H}_{\mathrm{so}}^{6\times 6}$ is diagonal in the basis of the total angular momentum eigenstates $|J, J_3\rangle$. This basis can be used to classify the valence band into heavy hole (hh), light hole (lh), and split-off hole (so) as follows

$$\begin{aligned} |\mathrm{hh}\uparrow\rangle = \left|\frac{3}{2}, \frac{3}{2}\right\rangle = -\frac{1}{\sqrt{2}}\left(|x_1\uparrow\rangle + i\,|x_2\uparrow\rangle\right), \\ |\mathrm{hh}\downarrow\rangle = \left|\frac{3}{2}, -\frac{3}{2}\right\rangle = \frac{1}{\sqrt{2}}\left(|x_1\downarrow\rangle - i\,|x_2\downarrow\rangle\right), \end{aligned} \tag{1.22}$$

$$
\begin{aligned}
\left|\mathrm{lh}\uparrow\right\rangle &= \left|\frac{3}{2},\frac{1}{2}\right\rangle = -\frac{1}{\sqrt{6}}\left(\left|x_1\downarrow\right\rangle + i\left|x_2\downarrow\right\rangle - 2\left|x_3\uparrow\right\rangle\right),\\
\left|\mathrm{lh}\downarrow\right\rangle &= \left|\frac{3}{2},-\frac{1}{2}\right\rangle = \frac{1}{\sqrt{6}}\left(\left|x_1\uparrow\right\rangle - i\left|x_2\uparrow\right\rangle + 2\left|x_3\downarrow\right\rangle\right),
\end{aligned}
\tag{1.23}
$$

$$
\begin{aligned}
\left|\mathrm{so}\uparrow\right\rangle &= \left|\frac{1}{2},\frac{1}{2}\right\rangle = \frac{1}{\sqrt{3}}\left(\left|x_1\downarrow\right\rangle + i\left|x_2\downarrow\right\rangle + \left|x_3\uparrow\right\rangle\right),\\
\left|\mathrm{so}\downarrow\right\rangle &= \left|\frac{1}{2},-\frac{1}{2}\right\rangle = \frac{1}{\sqrt{3}}\left(\left|x_1\uparrow\right\rangle - i\left|x_2\uparrow\right\rangle - \left|x_3\downarrow\right\rangle\right).
\end{aligned}
\tag{1.24}
$$

The spin-orbit splitting Δ_0 partly removes the degeneracy of the valence bands. At the band edge, their eigenenergies are given by

$$
E_{\mathrm{hh}}\left(\mathbf{k}=0\right) = E_{\mathrm{lh}}\left(\mathbf{k}=0\right) = E_v + \frac{\Delta_0}{3},\ E_{\mathrm{so}}\left(\mathbf{k}=0\right) = E_v - \frac{2\Delta_0}{3}. \tag{1.25}
$$

An even more accurate $\mathbf{k}\cdot\mathbf{p}$ model (especially for narrow gap semiconductors where nonparabolicities are important) can be obtained by enlarging our basis to include the lowest conduction-band s-orbital Bloch functions $\left|s\uparrow\right\rangle$ and $\left|s\downarrow\right\rangle$. The resulting eight-band Hamiltonian reads

$$
\hat{H}^{8\times8}\left(\mathbf{k}\right) = \begin{pmatrix} \hat{H}^{2\times2}_{\mathrm{cc}}\left(\mathbf{k}\right) & \hat{H}^{2\times6}_{\mathrm{cv}}\left(\mathbf{k}\right) \\ \hat{H}^{6\times2}_{\mathrm{vc}}\left(\mathbf{k}\right) & \hat{H}^{6\times6}_{\mathrm{vv}}\left(\mathbf{k}\right) \end{pmatrix}, \tag{1.26}
$$

with

$$
\hat{H}^{2\times2}_{\mathrm{cc}}\left(\mathbf{k}\right) = \left(E_c + \sum_{i=1}^{3} k_i A_c k_i\right) \otimes 1^{2\times2}, \tag{1.27}
$$

$$
\hat{H}^{2\times6}_{\mathrm{cv}}\left(\mathbf{k}\right) = \begin{pmatrix} \mathrm{i}Pk_1 + k_2Bk_3 & \mathrm{i}Pk_2 + k_3Bk_1 & \mathrm{i}Pk_3 + k_1Bk_2 \end{pmatrix} \otimes 1^{2\times2}, \tag{1.28}
$$

and

$$
\hat{H}^{6\times2}_{\mathrm{vc}}\left(\mathbf{k}\right) = \begin{pmatrix} -\mathrm{i}k_1P + k_3Bk_2 \\ -\mathrm{i}k_2P + k_1Bk_3 \\ -\mathrm{i}k_3P + k_2Bk_1 \end{pmatrix} \otimes 1^{2\times2}. \tag{1.29}
$$

Here, A_c includes the free-electron and remote band contributions to the conduction-band mass [19]. P denotes Kane's interband coupling matrix element. The parameter B is non-zero for crystals without inversion symmetry. Although, only for silicon and germanium $B = 0$ holds exactly, the parameter is usually neglected also for the III-V compounds. The valence-band part $\hat{H}^{6\times6}_{\mathrm{vv}}\left(\mathbf{k}\right)$ corresponds to Eq. (1.17) with only two modifications of parameters

$$
L \to L' = L + \frac{P^2}{E_g},\ N_+ \to N'_+ = N_+ + \frac{P^2}{E_g}. \tag{1.30}
$$

Due to the direct inclusion of conduction-band coupling, these modifications become necessary in order to avoid double counting of terms. Since the s-type conduction bands do not obtain any spin-orbit coupling, the eight-band spin-orbit Hamiltonian $\hat{H}_{\mathrm{so}}^{8\times8}$ is completely determined by the six-band case

$$\hat{H}_{\mathrm{so}}^{8\times8} = \begin{pmatrix} 0 & 0 \\ 0 & \hat{H}_{\mathrm{so}}^{6\times6} \end{pmatrix}. \tag{1.31}$$

In strained semiconductors, we add a strain-dependent Hamiltonian $\hat{H}_{\varepsilon}^{8\times8}$ which will be detailed in Sec. 1.4.2.

We note that the Hamiltonian (1.26) contains a particular ordering of the momentum operators that one might be tempted to simplify according to

$$k_i N'_+ k_j + k_j N_- k_i \overset{?}{=} N' k_i k_j, \text{ with } N' = N'_+ + N_-. \tag{1.32}$$

However, such a simplification only holds when the operators k_i and the material parameters commute. In the present case, this precise choice of operator ordering is required for the following reasons. First, in Refs. [15–17, 19], it has been shown that the specific ordering of Eq. (1.26) is required to obtain correct boundary conditions in the envelope function approximation where material parameters become position dependent (see next subsection). Second, a key element in the present work are magnetic fields $\mathbf{B}$. They can be introduced into the Hamiltonian (1.26) by replacing the kinetic momentum $\mathbf{k}$ with the canonic momentum $\mathbf{K} = \mathbf{k} + (e/\hbar)\,\mathbf{A}$ so that the following commutator relations hold

$$[K_i, K_j] = K_i K_j - K_j K_i = -\mathrm{i}\frac{e}{\hbar}\varepsilon_{ijk} B_k, \tag{1.33}$$

where ε_{ijk} denote the elements of the totally antisymmetric tensor. In Sec. 4.2.2, it will be shown that only the specific ordering of Eq. (1.26) guarantees the correct $\mathbf{k}{\cdot}\mathbf{p}$ Hamiltonian for bulk in nonzero magnetic fields [12].

Since we include the spin degree of freedom, finally we need to augment the Hamiltonian (1.26) by the Zeeman term,

$$\hat{H}_{\boldsymbol{\sigma}\cdot\mathbf{B}}^{8\times8} = \frac{\mathrm{g}_0 \mu_B}{2} \sum_{i=1}^{3} \hat{S}^i B_i, \tag{1.34}$$

where $\mathrm{g}_0 = 2$, and the matrix elements of the spin matrices $\hat{S}^i$ are given in terms of the Pauli matrices $\hat{\sigma}^i$,

$$S^i_{q'\sigma',q\sigma} = \langle q'\sigma' | \, \hat{\sigma}^i \, | q\sigma \rangle = \delta_{q',q} \hat{\sigma}^i_{\sigma',\sigma}. \tag{1.35}$$

So in total, the eight band $\mathbf{k}{\cdot}\mathbf{p}$ Hamiltonian is given by

$$\hat{H}_{\mathrm{k\cdot p}}^{8\times8}(\mathbf{k}) = \hat{H}^{8\times8}(\mathbf{k}) + \hat{H}_{\sigma B}^{8\times8} + \hat{H}_{\varepsilon}^{8\times8}, \tag{1.36}$$

with the individual contributions defined in Eqs. (1.26), (1.34), and (1.93).

Envelope function approximation

Up to now, we have considered only perfectly periodic bulk semiconductor crystals. In nanostructures, this periodicity is perturbed by a mesoscopic external potential $V(\mathbf{x})$. Such a potential can be created by local changes of band-edge energies due to the variation of material composition, local excess charges, and external electric fields. As a consequence, the translation invariance of the Hamiltonian is broken and Bloch's theorem does not hold for the wave functions. So the wave vector $\mathbf{k}$ is no good quantum number anymore and it is reasonable to switch to position space. In the envelope function approximation (EFA) [11], the wave function is written as a superposition of the Bloch functions at the extremum $\mathbf{k}_0$

$$\psi(\mathbf{x}) = \sum_n F^{(n)}(\mathbf{x})\, \psi_{n,\mathbf{k}_0}(\mathbf{x}), \tag{1.37}$$

where the expansion coefficients $F^{(n)}(\mathbf{x})$ are called envelope functions and are assumed to be slowly varying on the scale of the lattice constant. By this, most of the bulk $\mathbf{k}{\cdot}\mathbf{p}$ theory can be adopted for nanostructures. Namely, it can be shown [11] that for a homogenous semiconductor in an external potential, it holds a Schrödinger-like equation for the envelope functions

$$\sum_\mu \hat{H}^{\mathrm{EFA}}_{\nu\mu}(\mathbf{x})\, F^{(\mu)}(\mathbf{x}) = E F^{(\nu)}(\mathbf{x}), \tag{1.38}$$

with a Hamiltonian that corresponds to the bulk $\mathbf{k}{\cdot}\mathbf{p}$ Hamiltonian (1.13) Fourier transformed into real space ($\mathbf{k} \to -\mathrm{i}\boldsymbol{\nabla}$) plus the external potential

$$\hat{H}^{\mathrm{EFA}}_{\nu\mu}(\mathbf{x}) = \hat{H}^{\mathrm{k\cdot p}}_{\nu\mu}(-\mathrm{i}\boldsymbol{\nabla}) + \delta_{\nu\mu} V(\mathbf{x}). \tag{1.39}$$

For a nanostructure, the basic idea of this method is to patch up the bulk $\mathbf{k}{\cdot}\mathbf{p}$ Hamiltonian of each constituent material to obtain a global Hamiltonian for the envelope functions. In this case, all the $\mathbf{k}{\cdot}\mathbf{p}$ parameters become position dependent quantities and together with the momentum operators, a particular ordering has to be chosen (see also Sec. 1.2.2). In all Hamiltonians of the last subsection, the employed operator ordering (following Refs. [15–17, 19]) has been indicated; e.g. in Eq. (1.14), we have

$$k_1 N_+ k_2 \to -\frac{\partial}{\partial x_1} N_+(\mathbf{x}) \frac{\partial}{\partial x_2}. \tag{1.40}$$

The material parameters used in the present work have been taken from Refs. [22, 24].

We would like to note that the standard prescription of the EFA [11], is strictly speaking only valid for a homogenous semiconductor with slowly varying band edges $E_n(\mathbf{x})$ or external potentials $V(\mathbf{x})$. However, the method has been extensively applied to nanostructures and it has been shown that even in the presence of material induced discontinuities in the band edges, the EFA still delivers physically reasonable results. In addition, this empirical observation can also be justified using Burt's exact envelope function theory [15, 16].

Many nanodevices are structured only in one or two of the three spatial dimensions and still translation invariant in the remaining directions. For such devices, it is possible to separate the free motion of the charge carriers in the directions of translational invariance and the simulation domain can be reduced in dimension. In general, we call a device quasi d-dimensional if d is the number of structured directions. In this case, the EFA Schrödinger equation [Eq. (1.38)] reads

$$\sum_{\mu} \hat{H}^{\mathrm{EFA}}_{\nu\mu}\left(\mathbf{x},\mathbf{k}_{\|}\right) F^{(\mu)}\left(\mathbf{x},\mathbf{k}_{\|}\right) = E\left(\mathbf{k}_{\|}\right) F^{(\nu)}\left(\mathbf{x},\mathbf{k}_{\|}\right), \tag{1.41}$$

where $\mathbf{x} = (x_1, ..., x_d)$ is a d-dimensional position vector and $\mathbf{k}_{\|} = (k_{d+1}, ..., k_3)$ is a $(3-d)$-dimensional reciprocal lattice wave vector, with $1 \leq d \leq 3$. The Hamiltonian $\hat{H}^{\mathrm{EFA}}_{\nu\mu}\left(\mathbf{x},\mathbf{k}_{\|}\right)$ corresponds to the one from Eq. (1.39), where the $\mathbf{k}_{\|}$-directions have not been Fourier transformed into real space. Note that for $d = 3$, $\mathbf{k}_{\|}$ becomes useless and Eq. (1.41) reduces to Eq. (1.38).

The solution of the Schrödinger equation requires the specification of its boundary conditions. At the boundary $\partial\Omega$ of the simulation domain Ω, we either employ Dirichlet-, von Neumann-, or periodic boundary conditions. In general, these conditions are given by

$$\psi\left(\mathbf{x}\right)\big|_{\mathbf{x}\in\partial\boldsymbol{\Omega}} = f\left(\mathbf{x}\right) \qquad \text{(Dirichlet boundary condition)}, \tag{1.42}$$

$$\partial_{\mathbf{n}}\psi\left(\mathbf{x}\right)\big|_{\mathbf{x}\in\partial\boldsymbol{\Omega}} = g\left(\mathbf{x}\right) \qquad \text{(von Neumann boundary condition)}, \tag{1.43}$$

$$\psi\left(\mathbf{x}\right)\big|_{\mathbf{x}\in\partial\boldsymbol{\Omega}^{(i)}_{L}} = \psi\left(\mathbf{x}\right)\big|_{\mathbf{x}\in\partial\boldsymbol{\Omega}^{(i)}_{R}} \qquad \text{(periodic boundary condition)}, \tag{1.44}$$

where $\mathbf{n}$ denotes the normal to $\partial\Omega$ and $f\left(\mathbf{x}\right)$, $g\left(\mathbf{x}\right)$ are given functions on the boundary of Ω (taken to be zero for the Schrödinger equation). In the periodic boundary condition, the wave functions are set to be equal at the left and right boundaries in the periodic i-direction $\partial\boldsymbol{\Omega}^{(i)}_{L}$, and $\partial\boldsymbol{\Omega}^{(i)}_{R}$, respectively. Here, each element in $\partial\boldsymbol{\Omega}^{(i)}_{R}$ can be obtained by translating an element in $\partial\boldsymbol{\Omega}^{(i)}_{L}$ by $L_i\mathbf{e}_i$, where L_i is the length of the simulation domain in the direction of the unit vector $\mathbf{e}_i$. This boundary condition is needed for periodically repeated structures such as superlattices.

In order to actually calculate the eigenstates of the EFA Hamiltonian, Eq. (1.41) is transformed into a matrix eigenvalue problem by mapping it onto a discrete real-space grid. This discretization will be discussed in Sec. 2.3, and the solution algorithms for the resulting matrix eigenvalue problem are detailed in Sec. 2.5. For nonzero magnetic fields, the discretization is surprisingly problematic and will be discussed in chapter 3.

1.2.2 Spurious solutions

Even though the envelope function approach is well established, it is less widely known that this method is plagued by several ambiguities and instabilities that can lead to unphysical ghost states [19, 26], incorrect bound states at interfaces [27], and artificial oscillatory wave functions. We have been able to eliminate these problems by a careful treatment of so-called remote-band contributions to the bulk **k·p** Hamiltonians, by defining an operator ordering that leads to manifestly self-adjoint Hamiltonians, and by employing an upwinding scheme for the discretization of derivatives. We will only briefly sketch the rationale behind these techniques and refer the reader to Ref. [28] for a more detailed discussion.

Large wave vector solutions

In principle, spurious solutions can come with any multiband **k·p** Hamiltonian. For a qualitative discussion, we use a simple one-dimensional two-band model [29] that already includes all of the problems that may occur

$$\hat{H}^{2\times2}(k) = \begin{pmatrix} E_c + A_c k^2 & \mathrm{i}Pk \\ -\mathrm{i}Pk & E_v + A_v k^2 \end{pmatrix}. \tag{1.45}$$

This Hamiltonian couples a single conduction band (c) with a single valence band (v) via the interband momentum matrix element P. The parameters A_c and A_v describe free-electron and remote-band contributions to the effective masses. The secular equation $\det[\hat{H}^{2\times2}(k) - EI^{2\times2}] = 0$ of this Hamiltonian contains a term $A_c A_v k^4$ and therefore exhibits two solutions of k^2 for any value of E. Only one of these solutions may be physical, since conduction- and valence band do not overlap in a semiconductor. For $k^2 < 0$ ($A_c A_v < 0$), the additional solutions are evanescent [29] and pose no actual problem [28]. For $k^2 > 0$ ($A_c A_v > 0$), however, they form oscillatory modes [26] that can lead to spurious ghost states. This is a realistic scenario for the eight-band model of Eq. (1.26), which can be reduced to the present Hamiltonian (1.45) for $k \| [001]$ neglecting spin and decoupled bands. Using the material parameters from Ref. [22] for InAs, we then have $A_c = -4.8\,\hbar^2/(2m_0)$ and $A_v = L' + \hbar^2/(2m_0) = -14.7\,\hbar^2/(2m_0)$. In Fig. 1.2(a), a band structure is shown that corresponds to this situation. The conduction band $E_c(k)$ has small energy solutions also at large values of k, and even worse, the band structure does not possess a band gap. In a bulk semiconductor, this does not pose a problem since one can reject all large-k solutions that are beyond the validity of the **k·p** model, anyway. In a superlattice or heterostructure, however, an artificial small Brillouin zone is introduced. The large k values are folded back to small ones and unphysical ghost states are created. In Fig. 1.2(b), we show calculated eigenstates obtained by augmenting the Hamiltonian of Fig. 1.2(a) with a locally varying potential that produces a 10 nm quantum well. The spurious states can either be strongly oscillating states energetically lying within the band gap, or bound quantum well states that are modulated by an unphysical high frequency component. Following Ref. [19], we can avoid this problem by setting $A_c = 0$ and using the parameter P to fit the effective mass of the

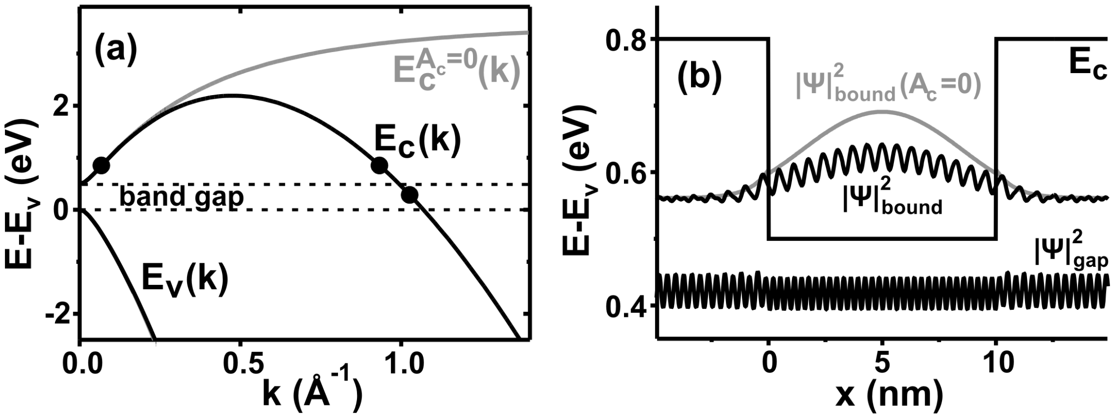


Figure 1.2: (a) Band structure of two-band Hamiltonian (1.45) for $A_c, A_v < 0$ (black lines) and $A_c = 0, A_v < 0$ (gray line). Circles indicate eigenenergies of appropriate quantum well eigenstates. (b) Conduction band edge of quantum well and probability densities of eigenstates, shifted to their eigenenergies. The solution for $A_c = 0$ has been rescaled for better visualization.

conduction band m_c^* according to the relation

$$\frac{1}{m_c^*} = A_c + P^2 \frac{E_g + 2/3\Delta}{E_g\left(E_g + \Delta\right)} \equiv \widetilde{P}^2 \frac{E_g + 2/3\Delta}{E_g\left(E_g + \Delta\right)}. \tag{1.46}$$

In the eight-band model, this requires to rescale certain valence band parameters analogous to Eq. (1.30). Due to this change of parameters, the k^4 term is removed from the dispersion and only the correct small wave vector solutions remain. The resulting modification of the band structure and the position space eigenfunctions is shown in Fig. 1.2. Alternatively setting $A_v = 0$ is unfavorable, because in three dimensions the valence band mass is anisotropic and can not be fitted by a single parameter. Another approach that is often used is to set $A_c = \hbar^2/\left(2m_0\right)$ which corresponds to entirely neglecting remote-band contributions. By this, the spurious ghost states are removed in most cases (since $A_v > 0$ hardly occurs) and only the uncritical evanescent modes remain.

Operator ordering

Having provided a solution for the problems related with the bulk band structure, additional issues occur when advancing to heterostructures. Here, one must also determine the proper arrangement of differential operators with respect to position dependent material parameters. E.g. in Eq. (1.45), we have a term iPk with an ordering that is not known a priori when transformed into real space

$$\mathrm{i}Pk \overset{?}{\rightarrow} \begin{cases} \frac{d}{dx}P\left(x\right) \\ P\left(x\right)\frac{d}{dx} \end{cases}. \tag{1.47}$$

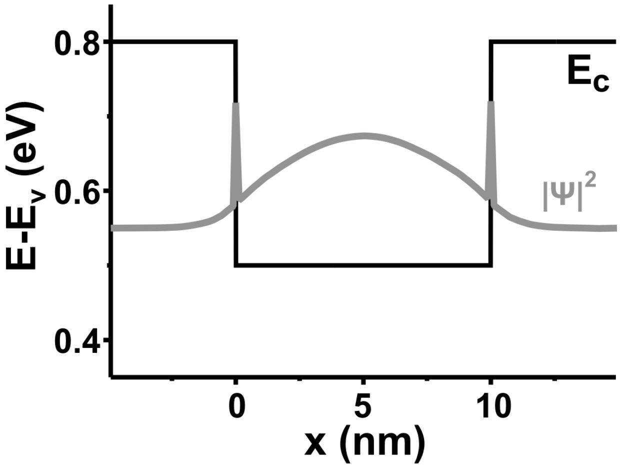


Figure 1.3: Conduction band edge of quantum well and probability density of bound state with interface spikes that has been calculated using the naively symmetrized Hamitonian (1.48) with the parameters discussed in the main text.

While it is clear that any operator ordering must yield a Hermitian Hamiltonian $\hat{H}_{\nu\mu}(x) = \hat{H}^\dagger_{\mu\nu}(x)$, this requirement is not sufficient to uniquely define $\hat{H}_{\nu\mu}(x)$. The customary technique of symmetrizing each matrix element of the Hamiltonian (i.e. requiring $\hat{H}_{\nu\mu}(x) = \hat{H}^\dagger_{\nu\mu}(x)$) has never been justified. In fact, such a Hamiltonian can lead to inconsistent boundary conditions for the envelope functions at material interfaces [19]. This corresponds to a Hamiltonian that is Hermitian but not self-adjoint [27]. For a finite dimensional Hamiltonian that is obtained by discretization of differential operators on a real-space grid, certain artifacts of this inconsistency can remain. In Fig. 1.3, we show a bound quantum well state that has been obtained by solving the naively symmetrized Hamiltonian

$$\hat{H}_{\text{sym}}(x) = \begin{pmatrix} E_c(x) & \frac{1}{2}\left[\frac{d}{dx}P(x) + P(x)\frac{d}{dx}\right] \\ \frac{1}{2}\left[\frac{d}{dx}P(x) + P(x)\frac{d}{dx}\right] & E_v(x) - \frac{d}{dx}A_v(x)\frac{d}{dx} \end{pmatrix}, \tag{1.48}$$

on a finite grid. The wave function exhibits pronounced unphysical singularities at the material interfaces. Note that the size of these spikes actually depends on the grid spacing and the local variation of the P-parameter. In Fig. 1.3, we have used an extremely fine grid spacing of $\varepsilon = 0.0125$ nm to elucidate the problem, while keeping with realistic material parameters of $P_{\text{well}} = 0.845$ eVnm, $P_{\text{barrier}} = 0.67$ eVnm, $A_v = -6\,\hbar^2/(2m_0)$, and $E_g = 0.5$ eV. Alternatively using the asymmetric ordering proposed in Ref. [19]

$$\hat{H}_{\text{correct}}(x) = \begin{pmatrix} E_c(x) & P(x)\frac{d}{dx} \\ \frac{d}{dx}P(x) & E_v(x) - \frac{d}{dx}A_v(x)\frac{d}{dx} \end{pmatrix}, \tag{1.49}$$

yields wave functions that do not contain such deficiencies. We use this asymmetric ordering also for the k-linear terms in the three-dimensional eight-band Hamiltonian

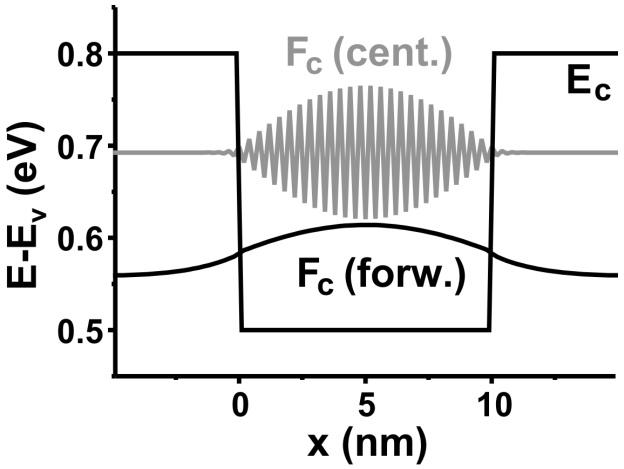


Figure 1.4: Conduction band edge of quantum well and conduction band amplitudes of bound states obtained by forward/backward differences (black) or centered differences (gray). The same parameters as in Fig. 1.3 have been used, except for the grid spacing that is now $\varepsilon = 0.2$ nm. The states have been shifted to different energies for better visualization.

(1.26). For the remaining terms, a unique operator ordering can be derived from Burt's exact envelope function theory [15–17]. This particular ordering has been indicated in all Hamiltonians of Sec. 1.2.1. Although, the individual elements are partly unsymmetric, the total Hamiltonians are Hermitian and self-adjoint and therefore retain a conservation of the probability flux density.

Numerical instabilities

A third issue that has not been discussed so far concerns numerical instabilities. When the envelope function Hamiltonian (1.49) is mapped onto a discrete real-space lattice, the discretization of the first order space derivatives is not unique. For example, one could use finite centered differences

$$\frac{d}{dx}F\left(x\right) \rightarrow \delta_x F\left(x\right)\big|_{\mathrm{cent}} = \frac{F\left(x+\varepsilon\right) - F\left(x-\varepsilon\right)}{2\varepsilon}, \tag{1.50}$$

where ε denotes the grid spacing. Unfortunately this leads to serious problems as unphysical solutions $F(x)$ that rapidly oscillate between nearest neighbor grid nodes with a relative phase of -1. In Fig. 1.4, we show the amplitude of the conduction band component of a bound quantum well state that has been obtained with the Hamiltonian (1.49) using centered finite differences. While the probability density of such a state in smooth, its amplitude is unphysically oscillating. The centered difference scheme is "blind" to these oscillations since it only compares values at second nearest neighbor points. In this approximation the oscillating wave function has a well defined small first derivative, although this is obviously not correct. Besides these incorrect states, also the correct smooth wave functions are obtained. For a homogenous material and the Hamiltonian (1.45), it can be shown that the incorrect conduction band states are shifted by $4A_c/\varepsilon^2$ with respect to the correct ones [28]. So for $A_c > 0$ and a sufficiently small grid spacing ε, they can be shifted out of the relevant part of the spectrum. For $A_c = 0$, however, they have to be removed explicitly. In order to eliminate this problem, we use forward or backward differencing to approximate first derivatives

$$\frac{d}{dx}F\left(x\right) \rightarrow \delta_x F\left(x\right)\big|_{\text{forw/backw}} = \pm\frac{F\left(x\pm\varepsilon\right) - F\left(x\right)}{\varepsilon}. \tag{1.51}$$

This approach is not compatible with the unphysically fast oscillating envelope functions since it compares values at nearest neighbor grid nodes (see Fig. 1.4). Since plain forward or backward differencing results into a non-Hermitian Hamilton matrix, we have employed a combination of forward and backward differencing that guarantees the Hermiticity of the Hamiltonian. Here, we use forward differencing for all first derivatives that are located in the blocks above the diagonal of the Hamiltonian and backward differencing for those below the diagonal,

$$\hat{H}_{\text{discrete}} = \begin{pmatrix} & \text{forward} \\ \text{backward} & \end{pmatrix}. \tag{1.52}$$

Note that in the diagonal blocks of our Hamiltonian, we do not have any first derivatives. With this discretization, a second-order derivative term is added to each first-order derivative that vanishes in the limit of zero grid spacing $\varepsilon \rightarrow 0$. This second-order term now suppresses the high-frequency components which constitute the spurious oscillating solutions, while the physically accurate low-frequency components remain unchanged. This approach is analogous to upwinding schemes that are used for hyperbolic differential equations in hydrodynamics [30]. More details on the discretization scheme will be given in Sec. 2.3.

1.3 External potential

In nanostructures which we describe in the envelope function approximation, the periodic semiconductor band structure is perturbed by a mesoscopic external potential $V\left(\mathbf{x}\right)$. Besides abrupt jumps in band-edge energies at material interfaces that have been stated in Sec. 1.2.1, $V\left(\mathbf{x}\right)$ also includes the electrostatic potential that results from local excess charges and external electric fields. In addition, we describe many particle effects by an exchange-correlation potential within a density functional approach. In the following, we will present the calculation of these individual contributions of the external potential.

1.3.1 Charge densities

We consider all charges in a nanostructure that deviate from the intrinsic charges of the semiconductor by a position dependent total charge density $\rho(\mathbf{x})$,

$$\rho(\mathbf{x}) = e\left[p(\mathbf{x}) - n(\mathbf{x}) + N_D^+(\mathbf{x}) - N_A^-(\mathbf{x}) + \rho_{\text{pol}}(\mathbf{x}) + \rho_{\text{fix}}(\mathbf{x})\right]. \tag{1.53}$$

This density will be included in the Hamiltonian by means of an electrostatic potential that can be obtained from the Poisson equation, as will be discussed in Sec. 1.3.2. The density $\rho(\mathbf{x})$ consists of free holes $p(\mathbf{x})$ and electrons $n(\mathbf{x})$, ionized donors $N_D^+(\mathbf{x})$ and acceptors $N_A^-(\mathbf{x})$, polarization charges $\rho_{\text{pol}}(\mathbf{x})$ and fixed charges $\rho_{\text{fix}}(\mathbf{x})$. The polarization charges will be given in Eq. (1.104) and result from strain induced piezo- and pyroelectric polarizations. Examples for fixed charges are surface and interface traps that come with lattice defects. They are not calculated but need to be specified explicitly.

For the calculation of free-carrier charges, we have employed two different models. In those parts of the nanostructures where quantum mechanical effects play an important role, we calculate charge densities by means of the eigenstates obtained from the multiband **k·p** envelope function method. In the remaining areas (typically lying at the boundaries of the simulation domain) where quantum mechanical effects can be neglected, we rely on the computationally less demanding classical Thomas-Fermi approximation.

Quantum mechanical charge densities

In a quasi d-dimensional nanostructure, the general *quantum mechanical* multiband charge densities of electrons and holes are given by

$$\begin{aligned} n_{\text{qm}}(\mathbf{x}) &= \sum_{i\in\text{CB}} \frac{g}{(2\pi)^{3-d}} \int_{\Omega_{BZ}} d^{3-d}\mathbf{k}_\parallel \left|F_i\left(\mathbf{x},\mathbf{k}_\parallel\right)\right|^2 f\left(\frac{-E_i\left(\mathbf{k}_\parallel\right) + E_{F,n}(\mathbf{x})}{k_B T}\right), \\ p_{\text{qm}}(\mathbf{x}) &= \sum_{i\in\text{VB}} \frac{g}{(2\pi)^{3-d}} \int_{\Omega_{BZ}} d^{3-d}\mathbf{k}_\parallel \left|F_i\left(\mathbf{x},\mathbf{k}_\parallel\right)\right|^2 f\left(\frac{E_i\left(\mathbf{k}_\parallel\right) - E_{F,p}(\mathbf{x})}{k_B T}\right). \end{aligned} \tag{1.54}$$

Here, the sums over i run over the eigenstates lying in the conduction- (CB) and valence bands (VB). The appropriate envelope functions F_i and eigenenergies E_i result from the EFA Schrödinger equation [Eq. (1.41)]. Note that the probability density for a given subband i in a N_A-band model is given by

$$\left|F_i\left(\mathbf{x},\mathbf{k}_\parallel\right)\right|^2 = \sum_{\nu=1}^{N_A} \left|F_i^{(\nu)}\left(\mathbf{x},\mathbf{k}_\parallel\right)\right|^2. \tag{1.55}$$

The factor g accounts for possible spin and valley degeneracies and depends on the particular **k·p** model. The maximum single-band degeneracies (multiband: $g = 1$)

are given by $g(\Gamma) = 2$, $g(X) = 6$, $g(L) = 8$, $g(\Delta) = 12$ and can be reduced by strain and magnetic fields (see also Sec. 1.2.1). All states are occupied according to the Fermi distribution function $f(E)$, with position dependent electron and hole quasi Fermi levels $E_{F,n}(\mathbf{x})$ and $E_{F,p}(\mathbf{x})$, respectively. These specific Fermi levels are used to model also non-equilibrium situations, as will be discussed in Sec. 1.5. We want to express that in semiconductor nanostructures, a situation can occur where a separate occupation of electron and hole states is not possible anymore. In chapter 6, we will present a novel method for such broken-gap structures.

For $d = 3$, the wave vectors in Eq. (1.54) are useless and the integral over $\mathbf{k}_{\|}$ can be ignored. For $d < 3$, the wave vectors are restricted to the $(3-d)$-dimensional Brillouin zone Ω_{BZ} in the reciprocal $\mathbf{k}_{\|}$ space. In this case, the charge densities are obtained by integrating over $\mathbf{k}_{\|}$. Instead of integrating over the full square area of the Brillouin zone Ω_{BZ}, in fact it is sufficient to integrate over the irreducible wedge Ω_{IW} of the Brillouin zone that can be obtained by exploiting the symmetry of the lattice [31]. After this reduction of computational effort is utilized, all densities have to be multiplied by the ratio Ω_{BZ}/Ω_{IW} to retain their original values. In addition, states that deviate from the Fermi energy by more than a few k_BT hardly contribute to the sums and integrals in Eq. (1.54), since the Fermi function drops exponentially for energies that are larger than the Fermi energy. Therefore, it usually suffices to integrate over Ω_{IW} only for small values of the wave vector modulus (typically less than 10% of the maximum k value in the bulk Brillouin zone). Similarly, higher energy states that do not significantly contribute to the charge density are not calculated at all. The remaining inner $\mathbf{k}_{\|}$ space is mapped onto a square wave vector lattice, where the Schrödinger equation is actually solved only for a discrete set of $\mathbf{k}_{\|}$ vectors. In order to limit the required amount of $\mathbf{k}_{\|}$ values, we have implemented an efficient k-space integration scheme similar to Refs. [32, 33] that interpolates the probability densities and the energy dispersion between the exactly solved $\mathbf{k}_{\|}$ points in order to guarantee well converged results. It turns out that for the probability density a bilinear (linear for $d = 2$) interpolation is sufficient, whereas for the energy dispersion higher order schemes are required to accurately sample energy minima with a large density of states. In fact, we rely on a $(3-d)$-dimensional cubic spline interpolation [34] for the energy dispersion. All the details of this procedure can be found in Ref. [6].

In the single-band models (EMA) and for $d < 3$, the integration over $\mathbf{k}_{\|}$ can be performed analytically due to the parabolic energy dispersion. This leads to the following simplifications for the electron and hole charge densities [Eq. (1.54)] that require to solve the Schrödinger equation only for $\mathbf{k}_{\|} = 0$

$$n_{\mathrm{qm}}^{\mathrm{EMA}}(\mathbf{x}) = \sum_{i\in\mathrm{CB}} g N_i^{(3-d)}(T)\,|F_i(\mathbf{x},0)|^2\,\mathcal{F}_{(1-d)/2}\left(\frac{-E_i(0)+E_{F,n}(\mathbf{x})}{k_BT}\right),$$

$$p_{\mathrm{qm}}^{\mathrm{EMA}}(\mathbf{x}) = \sum_{i\in\mathrm{VB}} g N_i^{(3-d)}(T)\,|F_i(\mathbf{x},0)|^2\,\mathcal{F}_{(1-d)/2}\left(\frac{E_i(0)-E_{F,p}(\mathbf{x})}{k_BT}\right). \qquad (1.56)$$

Here, the function $\mathcal{F}_n(E)$ denotes the Fermi-Dirac integral of order n

$$\mathcal{F}_n(E) = \frac{1}{\Gamma(n+1)} \int_0^\infty d\varepsilon \, \frac{\varepsilon^n}{\exp(\varepsilon - E) + 1}, \tag{1.57}$$

which is evaluated numerically [35]. The integration over $\mathbf{k}_\parallel$ results in the effective density of states of a $(3-d)$-dimensional free electron gas

$$N_i^{(3-d)}(T) = \left(\frac{m_i^{\mathrm{DOS}} k_B T}{2\pi\hbar^2} \right)^{(3-d)/2}, \tag{1.58}$$

with the density of state mass m_i^{DOS} of the i-th subband of the Hamiltonian. Since in a nanostructure we have a position dependent effective mass tensor $\hat{m}^*(\mathbf{x})$, a straight-forward determination of m_i^{DOS} from $\hat{m}^*(\mathbf{x})$ would lead to unphysical abrupt changes in the charge densities at material interfaces. In order to obtain a position independent density of state mass, we weight the effective mass tensor with the probability density of the corresponding subband according to the following relation

$$m_i^{\mathrm{DOS}} = \int_\Omega d^d\mathbf{x} \, |F_i(\mathbf{x}, 0)|^2 \left[\det \hat{m}^*_{(3-d)\times(3-d)}(\mathbf{x}) \right]^{1/(3-d)}. \tag{1.59}$$

Here, $\hat{m}^*_{(3-d)\times(3-d)}(\mathbf{x})$ denotes the $(3-d) \times (3-d)$ submatrix of the effective mass tensor $\hat{m}^*(\mathbf{x})$ [Eq. (1.14)] in the $\mathbf{k}_\parallel$-space. I.e. for $d = 1$, we have

$$\hat{m}^*_{2\times 2}(\mathbf{x}) = \begin{pmatrix} m^*_{2,2}(\mathbf{x}) & m^*_{2,3}(\mathbf{x}) \\ m^*_{3,2}(\mathbf{x}) & m^*_{3,3}(\mathbf{x}) \end{pmatrix}, \tag{1.60}$$

while for $d = 2$, the submatrix reduces to the scalar element $m^*_{3,3}(\mathbf{x})$ and for $d = 3$, the integration over $\mathbf{k}_\parallel$ is not needed at all. The integral in Eq. (1.59) extends over the spatial area Ω where the Schrödinger equation is solved.

It turns out that sometimes it is useful to apply different $\mathbf{k}\cdot\mathbf{p}$ models simultaneously. E.g. when using the eight-band model for the Γ conduction- and valence bands, the X and L conduction band valleys (which we employ in the EMA model) might contribute to the charge densities as well. In this situation, we sum up the charge densities that result from Eqs. (1.54) or (1.56) for the individual EFA-Hamiltonians.

Classical charge densities

In areas where quantum mechanical effects do not play a role, we calculate free-carrier charge densities in the, *classical* Thomas-Fermi approximation. In this approach, the following expressions for the charge densities of electrons and holes can be derived

$$\begin{aligned} n_{\mathrm{cl}}(\mathbf{x}) &= \sum_{\mu\in\mathrm{CB}} g_\mu N_\mu^{\mathrm{3D}}(\mathbf{x}, T) \, \mathcal{F}_{1/2}\left(\frac{-E_\mu(\mathbf{x}) - V(\mathbf{x}) + E_{F,n}(\mathbf{x})}{k_B T} \right), \\ p_{\mathrm{cl}}(\mathbf{x}) &= \sum_{\mu\in\mathrm{VB}} g_\mu N_\mu^{\mathrm{3D}}(\mathbf{x}, T) \, \mathcal{F}_{1/2}\left(\frac{E_\mu(\mathbf{x}) + V(\mathbf{x}) - E_{F,p}(\mathbf{x})}{k_B T} \right). \end{aligned} \tag{1.61}$$

Here, the sums run over the relevant conduction- and valence band valleys (CB $= \{\Gamma, L, X, \Delta\}$, VB $= \{\mathrm{hh}, \mathrm{lh}, \mathrm{so}\}$). $E_\mu(\mathbf{x})$ denotes the position dependent band edge, $V(\mathbf{x})$ is the external potential and $N_\mu^{\mathrm{3D}}(\mathbf{x}, T)$ is the position dependent effective density of states of a 3-dimensional free electron gas

$$N_\mu^{\mathrm{3D}}(\mathbf{x}, T) = \left(\frac{m_\mu^{\mathrm{DOS}}(\mathbf{x})\, k_B T}{2\pi\hbar^2}\right)^{3/2}, \tag{1.62}$$

with the local density of state mass

$$m_\mu^{\mathrm{DOS}}(\mathbf{x}) = \left[\det \hat{m}_\mu^*(\mathbf{x})\right]^{1/3}, \tag{1.63}$$

where $\hat{m}_\mu^*(\mathbf{x})$ is the 3×3 effective mass tensor at the band edge μ.

For the densities of ionized shallow impurities, we always apply the Thomas-Fermi approximation. In this context, the concentrations of ionized acceptors and donors are given by

$$\begin{aligned} N_A^-(\mathbf{x}) &= \sum_{i \in \text{Acceptors}} \frac{N_A^{(i)}(\mathbf{x})}{1 + g_A^{(i)} \exp\left(\left[E_v(\mathbf{x}) + V(\mathbf{x}) + E_A^{(i)} - E_{F,p}(\mathbf{x})\right] / k_B T\right)}, \\ N_D^+(\mathbf{x}) &= \sum_{i \in \text{Donors}} \frac{N_D^{(i)}(\mathbf{x})}{1 + g_D^{(i)} \exp\left(\left[-E_c(\mathbf{x}) - V(\mathbf{x}) + E_D^{(i)} + E_{F,n}(\mathbf{x})\right] / k_B T\right)}, \end{aligned} \tag{1.64}$$

where the sums run over all included types of acceptors and donors. Each type of impurity is characterized by its ionization energy $E_{A/D}^{(i)}$ (absolute value, relative to the appropriate band edge $E_{v/c}$), its ground state level degeneracy $g_{A/D}^{(i)}$, and its spatial concentration $N_{A/D}^{(i)}(\mathbf{x})$. Usually, the impurity degeneracies of acceptors and donors are $g_A = 4$ and $g_D = 2$, respectively.

1.3.2 Poisson equation

The total charge density $\rho(\mathbf{x})$ [Eq. (1.53)] influences the global electronic structure via the electrostatic potential $\phi(\mathbf{x})$, which obeys the Poisson equation

$$\boldsymbol{\nabla} \hat{\varepsilon}(\mathbf{x})\, \boldsymbol{\nabla} \phi(\mathbf{x}) = -\rho(\mathbf{x}). \tag{1.65}$$

Here, $\hat{\varepsilon}(\mathbf{x})$ is in general a position dependent dielectric tensor and the resulting "built-in" electrostatic potential $\phi(\mathbf{x})$ is used to augment the envelope function Hamiltonian by the Hartree potential

$$V(\mathbf{x}) = V_H(\mathbf{x}) = -e\phi(\mathbf{x}). \tag{1.66}$$

In zincblende, the dielectric tensor reduces to a scalar ε, while in wurtzite it is given by $\hat{\varepsilon} = \varepsilon_a 1^{3\times3} + (\varepsilon_c - \varepsilon_a)\, \mathbf{e}_z \mathbf{e}_z^T$, with two independent components ε_a and ε_c.

The solution of the Poisson equation requires the specification of its boundary conditions. Here, we rely again on the Dirichlet-, von Neumann-, and periodic boundary conditions of Eqs. (1.42)-(1.44), where $\psi(\mathbf{x})$ now has to be replaced by $\phi(\mathbf{x})$. Analogous to the Schrödinger equation, the Poisson equation as well as the other partial differential equations in the following sections are mapped onto a discrete real-space grid (cf. Sec. 2.3). By this, the Poisson equation becomes a large system of linear equations. Since the total charge density $\rho(\mathbf{x})$, depends itself on the electrostatic potential $\phi(\mathbf{x})$ when the classical charge density [Eq. (1.61)] is used, the Poisson equation [Eq. (1.65)] may in fact become a nonlinear problem. The numerical solution of linear as well as nonlinear systems of equations will be discussed in Sec. 2.4. When quantum mechanical charge densities [Eq. (1.54)] are used, the Poisson equation is coupled with the Schrödinger equation and both equations need to be solved in a self-consistent way. An efficient solution strategy for this problem will be shown in Sec. 2.6.1.

External electric fields are incorporated into the calculation of nanostructures via electric contacts. These contacts are areas of the simulation domain, where the Poisson equation is not solved, but where the electrostatic potential is determined based on certain assumptions. The resulting $\phi(\mathbf{x})$ in the contacts is used to specify boundary conditions for the Poisson equation at the interfaces between the contacts and the remaining simulation area. E.g. in order to apply a constant electric field to a nanostructure, we can specify two contacts at opposite boundaries of the simulation domain, with different Dirichlet values ϕ_1 and $\phi_2 = \phi_1 + U$. Here, U corresponds to the applied bias. If the nanostructure does not contain any excess charges, the electrostatic potential resulting from Eq. (1.65) will change linearly between the two contacts. Then, the electric field which results from the electrostatic potential via

$$\mathbf{F}(\mathbf{x}) = -\boldsymbol{\nabla}\phi(\mathbf{x}), \tag{1.67}$$

is constant and directed along the connection between the two contacts. All implemented contact models will be discussed in detail in the context of the current calculation in Sec. 1.5. When no contacts are specified at all, we use von Neumann boundary conditions [Eq. 1.43)] with zero derivatives at all boundaries of the simulation domain. This corresponds to vanishing normal electric fields and therefore globally charge neutral devices.

1.3.3 Multi-particle effects

We describe multi-particle effects by an exchange-correlation potential within a density functional approach. The DFT method [36, 37] is based on the central assertion that for any interacting system of particles, a local single-particle potential $V_{\mathrm{ni}}(\mathbf{x})$ exists, such that the exact ground state density $n_0(x)$ equals the ground state density of the non-interacting problem $n_{\mathrm{ni}}(x)$. This non-interacting density can be obtained from the single-particle states of the non-interacting Hamiltonian. Therefore, with the exact knowledge of $V_{\mathrm{ni}}(\mathbf{x})$, the ground state energy and density of an interacting multi-particle state can be obtained by a self-consistent solution with multiple eigenstates from a single-particle Schrödinger equation.

In the present case, the non-interacting density $n_{\mathrm{ni}}(x) \equiv n(x)$ is obtained from the single-particle states of the EFA Hamiltonian (1.39) according to Eq. (1.54). Here, the potential $V_{\mathrm{ni}}(\mathbf{x}) \equiv V(\mathbf{x}) = V_H(\mathbf{x}) + V_{xc}(\mathbf{x})$ is composed of the Hartree term $V_H(\mathbf{x}) = -e\phi(\mathbf{x})$ (that is determined by the Poisson equation) and an exchange-correlation term $V_{xc}(\mathbf{x})$. The latter one is known exactly only for special cases. For general many-particle systems, we rely on the local spin density approximation (LSDA) [38]. In this scheme, the system is assumed to locally behave as a homogenous gas of charged carriers, for which exchange and correlation can be evaluated exactly as a function of the density. The LSDA is a refinement of the local density approximation (LDA) since it also accounts for the spin polarization $\zeta(\mathbf{x}) = [n_\uparrow(\mathbf{x}) - n_\downarrow(\mathbf{x})] / [n_\uparrow(\mathbf{x}) + n_\downarrow(\mathbf{x})]$, with spin-up and spin-down carrier densities $n_\uparrow(\mathbf{x})$ and $n_\downarrow(\mathbf{x})$, respectively. This is important especially for few-particle states and odd numbers of carriers. In this model, the exchange-correlation potential becomes spin dependent and whenever the single-band Schrödinger equation is used, the solutions become two-component spinors. The parametrization of the LSDA exchange-correlation potentials $V_{xc}^{\uparrow\downarrow}(\mathbf{x})$ and energies $\varepsilon_{xc}(\mathbf{x})$ is given in Appendix B. With this potential, the EFA Kohn-Sham-equation has to be solved self-consistently for all particles until the individual densities $n_\uparrow(\mathbf{x})$ and $n_\downarrow(\mathbf{x})$ become stationary. The total energy of the many-body state is finally obtained from

$$E_0 = \sum_i E_i + \frac{1}{2}\int_\Omega d^3x\, e\phi(\mathbf{x})\, n(\mathbf{x}) + \int_\Omega d^3x\, \varepsilon_{xc}(\mathbf{x}) n(\mathbf{x}) - \int_\Omega d^3x\, V_{xc}^{\uparrow\downarrow}(\mathbf{x}) n(\mathbf{x}), \quad (1.68)$$

where E_i are the eigenenergies of the single-particle states.

1.4 Elastic strain

Lattice mismatch between different semiconductor materials that are grown on top of each other results in displacements of the individual atoms from their positions of rest. This produces elastic strain that can strongly affect the electronic structure and therefore needs to be accounted for in any realistic calculation of semiconductor nanostructures. In the following, we will describe the calculation of strain effects in terms of linear continuum elasticity theory [39, 40].

1.4.1 Strain calculation

In a continuum model, a crystal can be described by a field of material points with coordinates $\mathbf{x}$. A distortion of the crystal shifts any point to a new position $\mathbf{x}' = \mathbf{x}'(\mathbf{x})$.

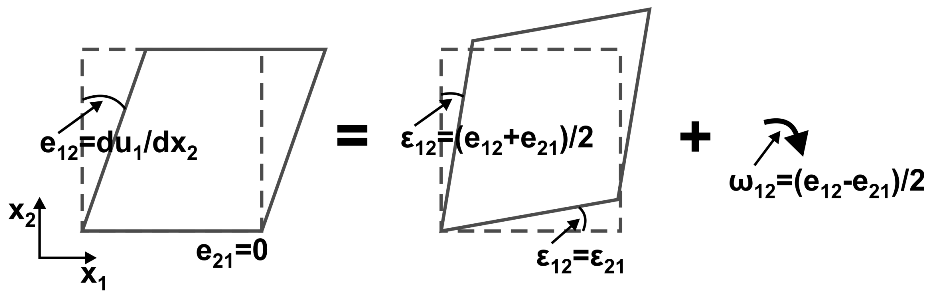


Figure 1.5: Illustration for off-diagonal elements of the distortion tensor. A general shear deformation (given by e_{ij}) consists of a pure shear deformation ε_{ij} and a rotation ω_{ij}.

This motivates the definition of a field of displacement vectors

$$\mathbf{u}(\mathbf{x}) = \mathbf{x}'(\mathbf{x}) - \mathbf{x}. \tag{1.69}$$

Since a uniform displacement of the whole crystal does not lead to any strain, we are only interested in nonuniform displacements. Such deformations can be described by a distortion tensor

$$e_{ij} = \frac{\partial u_i}{\partial x_j} \ (i, j = 1..3), \tag{1.70}$$

which can be split into a symmetric part that is named strain tensor

$$\varepsilon_{ij} = \frac{1}{2}(e_{ij} + e_{ji}), \tag{1.71}$$

and an asymmetric part

$$\omega_{ij} = \frac{1}{2}(e_{ij} - e_{ji}). \tag{1.72}$$

We assume all tensor elements to be small $e_{ij} \ll 1$. The diagonal elements of the distortion tensor attribute length changes $dx_i' = (1 + \varepsilon_{ii})\, dx_i$, its trace corresponds to a change of volume

$$dV' = (1+\varepsilon_{11})(1+\varepsilon_{22})(1+\varepsilon_{33})\, dV \approx (1+\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33})\, dV = [1 + Tr(\varepsilon)]\, dV, \tag{1.73}$$

and the off-diagonal elements $\varepsilon_{ij(i\neq j)}$ arise due to shear deformations of the crystal. The asymmetric part ω_{ij} only describes rotations of the crystal (see Fig. 1.5) and therefore does not contribute to the elastic energy. This energy is given by [41]

$$E = \int_V \frac{1}{2} C_{ijkl}\varepsilon_{ij}\varepsilon_{kl} dV \ \ (i, j, k, l = 1..3), \tag{1.74}$$

with the fourth rank elasticity tensor C_{ijkl}. To obtain a condition for mechanical equilibrium, the elastic energy density is minimized. This leads to the relation

$$\frac{\partial \sigma_{kl}}{\partial x_l} = 0. \tag{1.75}$$

In a generalized form that includes external volume forces $\mathbf{f}$, the relation reads

$$\frac{\partial \sigma_{kl}}{\partial x_l} = f_k. \tag{1.76}$$

Here, $\hat{\sigma}$ denotes the Cauchy stress tensor. In linear approximation, this stress tensor is related to the strain tensor $\hat{\varepsilon}$ by means of Hooke's law

$$\sigma_{kl} = C_{klij}\varepsilon_{ij}. \tag{1.77}$$

In order to solve the differential equation (1.76), we use the following boundary conditions. At the borders of the simulation domain, we either apply von Neumann conditions [Eq. (1.43)] that correspond to vanishing normal stress, or periodic boundary conditions [Eq. (1.44)] for periodic structures. A completely unstrained substrate area can be described by Dirichlet boundary conditions [Eq. (1.42)] that set the displacement to a constant value. At material interfaces, we assume pseudomorphic growth, i.e., we neglect any lattice defects and all epitaxially grown atoms have to be placed on the positions of the substrate lattice in lateral direction. To guarantee these conditions, we first displace all material points to match the substrate lattice (see step (1) in Fig. 1.6) [41]. This substrate will from now on serve as a reference lattice. The displacements produce a lattice mismatch strain $\hat{\varepsilon}^0$,

$$\varepsilon^0_{ij}(\mathbf{x}) = \delta_{ij}\frac{a^s_i - a_i(\mathbf{x})}{a_i(\mathbf{x})}. \tag{1.78}$$

Here, $a_i(\mathbf{x})$ denotes the local lattice constant in i-direction and a^s_i is the corresponding value for the substrate. Now, the strain tensor can be written as

$$\hat{\varepsilon} = \widetilde{\varepsilon} + \hat{\varepsilon}^0, \tag{1.79}$$

where $\widetilde{\varepsilon}$ only comprises the deformations relative to the reference lattice. The lattice mismatch leads to an increase of the elastic energy and therefore implies a counteracting force in growth direction. Minimization of the elastic energy now corresponds to a relaxation of the lattice constant in growth direction (see step (2) in Fig. 1.6). The relaxation can induce shear distortions (cf. Fig. 1.5) due to off-diagonal elements in the elasticity tensor that couple different elements of the displacement vector. Concretely, we need to solve Eq. (1.76) only for the relative part $\widetilde{\varepsilon}$ of the strain tensor. Here, the lattice mismatch induces an additional force on the right-hand side and thus the differential equation is modified according to

$$\begin{aligned}\frac{\partial \sigma_{kl}}{\partial x_l} &= \frac{\partial}{\partial x_l}C_{klij}\varepsilon_{ij} = f_k \\ &\rightarrow \frac{\partial}{\partial x_l}C_{klij}\widetilde{\varepsilon}_{ij} = f_k - \frac{\partial}{\partial x_l}C_{klij}\varepsilon^0_{ij}.\end{aligned} \tag{1.80}$$

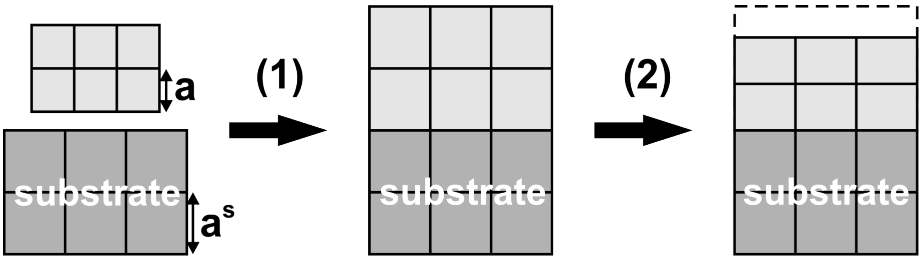


Figure 1.6: Illustration of strain calculation: (1) adjustment to substrate lattice, (2) relaxation by energy minimization.

Together with Eq. (1.71) and Eq. (1.78), this leads to

$$\frac{\partial}{\partial x_l} C_{klij} \frac{1}{2}\left(\frac{\partial \widetilde{u}_i}{\partial x_j} + \frac{\partial \widetilde{u}_j}{\partial x_i}\right) = f_k + \frac{\partial}{\partial x_l} C_{klij} \delta_{ij} \frac{a_i - a_i^s}{a_i}. \tag{1.81}$$

Utilizing the symmetry of the elasticity tensor

$$C_{ijkl} = C_{jikl} = C_{ijlk} = C_{klij}, \tag{1.82}$$

and by swapping the indices i and l, we finally obtain

$$\frac{\partial}{\partial x_i} C_{kilj} \frac{\partial}{\partial x_j} \widetilde{u}_l = f_k + \frac{\partial}{\partial x_i} C_{kill} \frac{a_l - a_l^s}{a_l}. \tag{1.83}$$

The solution of this differential equation determines the displacement vectors relative to the reference lattice $\widetilde{\mathbf{u}}(\mathbf{x})$ that can be used to calculate the total strain tensor,

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial \widetilde{u}_i}{\partial x_j} + \frac{\partial \widetilde{u}_j}{\partial x_i}\right) + \delta_{ij} \frac{a_i^s - a_i}{a_i}. \tag{1.84}$$

In heterostructures, analytic expressions can be found for this strain tensor (cf. Appendix C). The symmetry of the elasticity tensor accounts for the symmetry of the individual type of crystal. In order to write this fourth rank tensor as a matrix, one conventionally introduces the Voigt notation

$$11 \rightarrow 1, 22 \rightarrow 2, 33 \rightarrow 3, 23 \rightarrow 4, 13 \rightarrow 5, 12 \rightarrow 6. \tag{1.85}$$

In this notation, the six independent elements of the symmetric strain and stress tensors read

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \equiv \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ 2\varepsilon_{23} \\ 2\varepsilon_{13} \\ 2\varepsilon_{12} \end{pmatrix}, \quad \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} \equiv \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{pmatrix}. \tag{1.86}$$

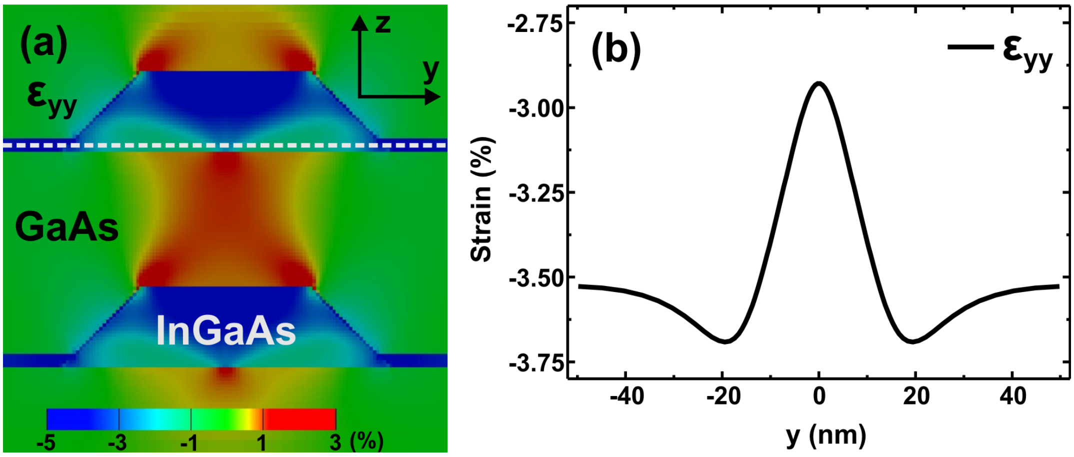


Figure 1.7: Cross sections of calculated strain tensor component ε_{yy} for vertically stacked InGaAs/GaAs double dot structure. (a) 2D cross section in yz-plane taken at the dot centers. (b) 1D cross section along the dashed white line in (a) when the upper dot is removed from the structure.

Finally, the elasticity tensor of zincblende and wurtzite crystals are given by [40]

$$C_{\mathrm{zb}} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & & & \\ C_{12} & C_{11} & C_{12} & & & \\ C_{12} & C_{12} & C_{11} & & & \\ & & & C_{44} & & \\ & & & & C_{44} & \\ & & & & & C_{44} \end{pmatrix}, \tag{1.87}$$

and [42]

$$C_{\mathrm{wz}} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & & & \\ C_{12} & C_{11} & C_{13} & & & \\ C_{13} & C_{13} & C_{33} & & & \\ & & & C_{44} & & \\ & & & & C_{44} & \\ & & & & & C_{66} \end{pmatrix}, \tag{1.88}$$

with $C_{66} = \frac{1}{2}\left(C_{11} - C_{12}\right)$.

To exemplify the importance of strain in semiconductor nanostructures, in Fig. 1.7, we present a calculated strain profile of a vertically stacked InGaAs/GaAs double dot structure that can form by self-assembly in epitaxial growth. The dots are assumed to have a width of 25 nm, a height of 8 nm, an interdot distance of 10 nm, a realistic trumpet-shaped alloy profile [43] and sit on 1 nm wetting layers with an indium content of 50%. Fig. 1.7(b) shows a cross section of the lateral strain tensor element ε_{yy} along a line in the upper wetting layer. This strain component is a measure for the relative change of the lattice constant in lateral direction. In contrast to Fig. 1.7(a), now the upper dot has been left out to model the situation in the growth process where the self-assembly of the upper dot actually starts. The graph shows a pronounced decrease of the strain modulus in the upper wetting layer at the center above the lower dot. This local reduction of lattice mismatch shows that it is energetically favorable for the upper dot to grow exactly above the lower one, as it has been observed experimentally [44]. Further properties of similar quantum dot molecules will be discussed in chapter 5.

In addition to growth processes, strain can strongly affect the electronic structure of semiconductors. First, it directly modifies the bulk band structure. Here, the dominant change is a shift of the band-edge energies that can be considered via deformation potentials. Second, it can produce polarization charges that modify the electrostatic potential and therefore have to be included in the Poisson equation. In the following, we will discuss these effects in some detail.

1.4.2 Deformation potentials

We consider strain induced shifts and splittings of band-edge energies via linear deformation potentials. The Hamiltonian (1.14) is augmented by an additional strain dependent part $\hat{H}_\varepsilon$ to take this contribution into account. For the conduction bands in zincblende materials, this term is given by [20]

$$\hat{H}_\varepsilon^\nu = \Xi_d^\nu Tr\left(\hat{\varepsilon}\right) + \Xi_u^\nu \left(\mathbf{e}_k \cdot \hat{\varepsilon} \cdot \mathbf{e}_k\right) \;\; \left(\nu = \Gamma, L, X, \Delta\right), \tag{1.89}$$

with

$$\Xi_d^\nu \equiv a_c^\nu - \frac{1}{3} a_u^\nu \text{ and } \Xi_u^\nu \equiv a_u^\nu. \tag{1.90}$$

Here, a_c^ν and a_u^ν denote the absolute and the uniaxial deformation potential of the $\nu \in \{\Gamma, L, X, \Delta\}$ conduction band, respectively. The unit vector $\mathbf{e}_k$ points in the direction of the respective minimum in k-space (Γ: $\mathbf{e}_k = 0$). Note that besides absolute energy shifts, Eq. (1.89) can also split the individual L-, X- and Δ-valleys. For the p-type valence band in the basis $\{\left|x\right\rangle, \left|y\right\rangle, \left|z\right\rangle\}$, the strain Hamiltonian is given by [9]

$$\hat{H}_\varepsilon^{3\times3} = \begin{pmatrix} l\varepsilon_{11} + m\left(\varepsilon_{22} + \varepsilon_{33}\right) & n\varepsilon_{12} & n\varepsilon_{13} \\ n\varepsilon_{12} & l\varepsilon_{22} + m\left(\varepsilon_{11} + \varepsilon_{33}\right) & n\varepsilon_{23} \\ n\varepsilon_{13} & n\varepsilon_{23} & l\varepsilon_{33} + m\left(\varepsilon_{11} + \varepsilon_{22}\right) \end{pmatrix}, \tag{1.91}$$

with

$$l = a_v + 2b, \;\; m = a_v - b, \;\; n = \sqrt{3}d. \tag{1.92}$$

Here, a_v is the absolute deformation potential and b as well as d are the shear deformation potentials of the valence band. In the eight-band Hamiltonian (1.36), the strain contribution is given by

$$\hat{H}_\varepsilon^{8\times8} = \begin{pmatrix} \hat{H}_\varepsilon^\Gamma \otimes 1^{2\times2} & 0 \\ 0 & \hat{H}_\varepsilon^{3\times3} \otimes 1^{2\times2} \end{pmatrix}. \tag{1.93}$$

Eq. (1.91) removes the degeneracy of light and heavy holes. E.g. in heterostructures that are grown in the [001] direction, i.e., for biaxial strain ($\varepsilon_{\|} = \varepsilon_{11} = \varepsilon_{22} \neq \varepsilon_{\perp} = \varepsilon_{33}$) without any shear deformations ($\varepsilon_{12} = \varepsilon_{13} = \varepsilon_{23} = 0$), the following energy shifts are produced [45]

$$\begin{aligned}
\Delta E_{\mathrm{HH}} &= a_v Tr\left(\varepsilon\right) - \delta E, & (1.94)\\
\Delta E_{\mathrm{LH}} &= a_v Tr\left(\varepsilon\right) + \frac{1}{2}\left(\delta E - \Delta_{so} + \sqrt{\Delta_{so}^2 + 2\Delta_{so}\delta E + 9\delta E^2}\right) & \\
&\approx a_v Tr\left(\varepsilon\right) + \delta E, & (1.95)\\
\Delta E_{\mathrm{SO}} &= a_v Tr\left(\varepsilon\right) + \frac{1}{2}\left(\delta E + \Delta_{so} - \sqrt{\Delta_{so}^2 + 2\Delta_{so}\delta E + 9\delta E^2}\right) & \\
&\approx a_v Tr\left(\varepsilon\right) - \frac{2\delta E^2}{\Delta_{so}}, & (1.96)
\end{aligned}$$

with $\delta E = b\left(\varepsilon_{\perp} - \varepsilon_{\|}\right)$. In the case of tensile strain ($\varepsilon_{\|} > 0$, $\varepsilon_{\perp} < 0$), the energy δE is positive, because the parameter b is negative for all relevant materials. Thus, the light hole is lifted above the heavy hole in this situation and vice versa for compressive strain. For general strain tensors, an analytical treatment of Eq. (1.91) becomes too complicated. Therefore, we calculate the band-edge shifts that are needed for the single-band Schrödinger equations of holes and the classical charge density [Eq. (1.61)] by diagonalizing $\hat{H}_{\varepsilon}$ at each position of the simulation domain, separately. In the multiband Hamiltonian, $\hat{H}_{\varepsilon}$ is included directly. In this way, also changes of effective masses can occur due to strain induced mixing of hole states. For wurtzite crystal structures, the strain contributions to the Hamiltonian can be found in Appendix A.

1.4.3 Polarization charges

In addition, strain in III-V semiconductors displaces the centers of positive and negative charges within the unit cells and therefore produces a so called piezoelectric polarization

$$P_i^{\mathrm{piezo}} = e_{ijk}\varepsilon_{jk} \ \left(i, j, k = 1..3\right), \tag{1.97}$$

with the material dependent piezoelectric tensor e_{ijk}. In zincblende crystals, this tensor has the following symmetry [46]

$$e_{ijk} = \begin{cases} e_{14} & \text{if } i \neq j \neq k \\ 0 & \text{else} \end{cases}. \tag{1.98}$$

Thus, only shear elements of the strain tensor contribute to the polarization

$$\mathbf{P}^{\mathrm{piezo}} = 2e_{14}\begin{pmatrix} \varepsilon_{23} \\ \varepsilon_{13} \\ \varepsilon_{12} \end{pmatrix}. \tag{1.99}$$

In wurtzite materials, also the diagonal elements of the strain tensor contribute to the piezoelectric polarization, because here we have [42]

$$e_{311} = e_{322} \equiv e_{31},\ e_{333} \equiv e_{33},\ e_{113} = e_{131} = e_{223} = e_{232} \equiv e_{15}, \tag{1.100}$$

and therefore

$$\mathbf{P}^{\text{piezo}} = \begin{pmatrix} 2e_{15}\varepsilon_{13} \\ 2e_{15}\varepsilon_{23} \\ e_{31}\varepsilon_{11} + e_{31}\varepsilon_{22} + e_{33}\varepsilon_{33} \end{pmatrix}. \tag{1.101}$$

At material interfaces where the piezoelectric constants change abruptly, or in an inhomogenous strain field, the microscopic electric dipoles are not compensated and produce mesoscopic dipole charges that can be calculated from the divergence of the polarization

$$\rho_{\text{piezo}} = -\operatorname{div}\mathbf{P}^{\text{piezo}} = -\partial_i e_{ijk}\varepsilon_{jk}. \tag{1.102}$$

These charges can strongly affect the electrostatic potential and therefore have to be included in the Poisson equation [Eq. (1.65)] via the total charge density [Eq. (1.53)]. Wurtzite materials are pyroelectric, i.e., they have a strain independent spontaneous polarization along the hexagonal c-axis

$$\mathbf{P}^{\text{pyro}} = P^{\text{pyro}}\mathbf{e}_3. \tag{1.103}$$

Analogous to Eq. (1.102), this leads to polarization charges at material interfaces that need to be taken into account

$$\rho_{\text{pol}} = \rho_{\text{piezo}} + \rho_{\text{pyro}}. \tag{1.104}$$

A typical situation, where piezoelectric charges have important consequences, are self-assembled quantum dots. In Fig. 1.8, we show the calculated piezoelectric charge distribution of an InAs quantum dot that is embedded in GaAs. The dot is assumed to have a realistic truncated pyramid shape. Due to the strong lattice mismatch and the specific geometry, we obtain large shear strains that lead to charge dipoles near the edges of the pyramid. In Sec. 5.3.2, we will investigate concrete consequences of these charges on the electronic structure of quantum dots.

1.5 Drift-diffusion current

When a nanostructure is subject to an external bias, the electronic system is driven out of equilibrium and electric currents flow. In this case, the Fermi levels differ between the individual contacts where the bias is applied. Thus, a single position independent Fermi level is not sufficient to describe the charge distribution, which in fact does not even follow the Fermi-Dirac statistics. Fully quantum mechanical calculations of carrier dynamics in quasi two- or three-dimensional devices are computationally still too demanding. So far, in next**nano**++ only the limiting cases of

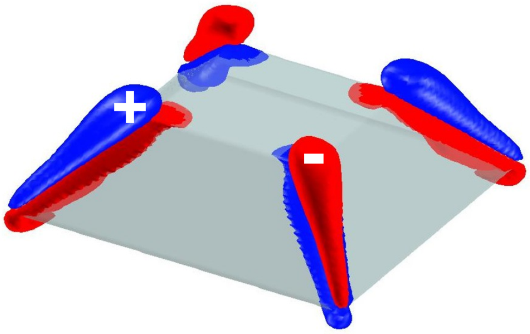


Figure 1.8: Isosurface of calculated piezoelectric polarization charge densities near the edges of an InAs quantum dot (light gray) that is embedded in GaAs (not shown). Positive and negative charges are indicated in blue and red, respectively.

highly diffusive or purely ballistic quantum mechanical transport can be calculated for such devices. While the ballistic transport model (CBR) has been detailed in Refs. [6, 7], we present here only the quantum drift-diffusion model (QDD).

This method extends the classical drift-diffusion method with a quantum mechanical calculation of the charge densities and has been introduced by Hackenbuchner [5]. Here, expressions for the continuity equation and the current density are derived from the lowest momenta of the Boltzman equation. Combining these expressions leads to the following set of equations that need to be solved

$$\begin{aligned} \boldsymbol{\nabla}\mu_n(\mathbf{x})\, n(\mathbf{x})\, \boldsymbol{\nabla}E_{F,n}(\mathbf{x}) &= R(\mathbf{x}), \\ \boldsymbol{\nabla}\mu_p(\mathbf{x})\, p(\mathbf{x})\, \boldsymbol{\nabla}E_{F,p}(\mathbf{x}) &= -R(\mathbf{x}). \end{aligned} \tag{1.105}$$

Here, $\mu_n(\mathbf{x})$ and $\mu_p(\mathbf{x})$ denote mobilities and $E_{F,n}(\mathbf{x})$ and $E_{F,p}(\mathbf{x})$ are locally varying quasi Fermi levels for electrons and holes, respectively. These Fermi levels are used to determine the charge densities $n(\mathbf{x})$ and $p(\mathbf{x})$ according to Eq. (1.54), assuming that the free-carrier charges are locally close to thermodynamic equilibrium and can be described by the Fermi statistics. The two equations are coupled in terms of generation and recombination processes that are both included in $R(\mathbf{x})$. Depending on the sign of R, either generation ($R < 0$) or recombination ($R > 0$) is dominant. In next**nano**++, we consider radiative (photon emitting), Schockley-Read-Hall (via deep traps), and Auger (three-particle processes) recombination models. A detailed description of their implementation together with the mobility models can be found in Appendix D of Ref. [6]. In fact, Eq. (1.105) is coupled with the Poisson equation [Eq. (1.65)] and the Schrödinger equation [Eq. (1.41)]. It is required to find solutions for the quasi Fermi levels, the electrostatic potential, and the eigenstates that fulfill all three equations simultaneously. The numerical solution of this task will be discussed in Sec. 2.6.2. Finally, when the quasi Fermi levels have been determined, the electron and hole current densities can be obtained from the relations

$$\mathbf{j}_n(\mathbf{x}) = -\mu_n(\mathbf{x})\, n(\mathbf{x})\, \boldsymbol{\nabla} E_{F,n}(\mathbf{x}),$$
$$\mathbf{j}_p(\mathbf{x}) = \mu_p(\mathbf{x})\, p(\mathbf{x})\, \boldsymbol{\nabla} E_{F,p}(\mathbf{x}). \quad (1.106)$$

A unique solution of the drift-diffusion current equation [Eq. (1.105)] requires the specification of boundary conditions. At the boundaries of the simulation domain, we always employ zero von Neumann boundary conditions [Eq. (1.43)], which implies vanishing current densities. Currents may only flow through the interfaces of contacts where we employ Dirichlet boundary conditions [Eq. (1.42)] to set the quasi Fermi levels to the applied bias

$$E_{F,n}(\mathbf{x})\big|_{\mathbf{x}\in\partial C} = E_{F,p}(\mathbf{x})\big|_{\mathbf{x}\in\partial C} = -e\phi_{\mathrm{bias}}. \quad (1.107)$$

Here, ∂C denotes the boundary surface of the contact. We would like to note that we have automated the integration of the current density over the contact surfaces to determine the total current flow. Some details of this procedure can be found in Appendix D. In next**nano**++, we have employed the following contact models that differ only in terms of the boundary condition of the Poisson equation:

- At an "Ohmic contact", we use von Neumann boundary conditions for the electrostatic potential

  $$\partial_{\mathbf{n}}\phi(\mathbf{x})\big|_{\mathbf{x}\in\partial C} = 0, \quad (1.108)$$

  which implies a vanishing electric field. This reflects the situation where the space charge region at the contact-semiconductor interface is negligible small so that charge carriers can tunnel unhampered. If a device contains only Ohmic contacts, global charge neutrality is guaranteed.

- At a "Schottky contact", we employ Dirichlet boundary conditions for the electrostatic potential

  $$\phi(\mathbf{x})\big|_{\mathbf{x}\in\partial C} = \frac{1}{e}E_c - \phi_{\mathrm{barrier}} + \phi_{\mathrm{bias}}. \quad (1.109)$$

  By this, the energetic difference between the conduction band edge E_c and the Fermi level is set to the Schottky barrier ϕ_{barrier}. This corresponds to the situation of Fermi level pinning due to surface states. The charge of the occupied surface states results in a non-vanishing electric field at the contact.

- At a "charge-neutral contact", we again employ Dirichlet boundary conditions for the electrostatic potential

  $$\phi(\mathbf{x})\big|_{\mathbf{x}\in\partial C} = \frac{1}{e}E_c - \phi_{\mathrm{neutral}} + \phi_{\mathrm{bias}}. \quad (1.110)$$

  Here, the Dirichlet value for the potential ϕ_{neutral} is determined by requiring local charge neutrality at the contact surface. This model is used to describe Ohmic contacts for devices that are too short or do not cover enough charges to obtain the full voltage drop in the "Ohmic contact" model.

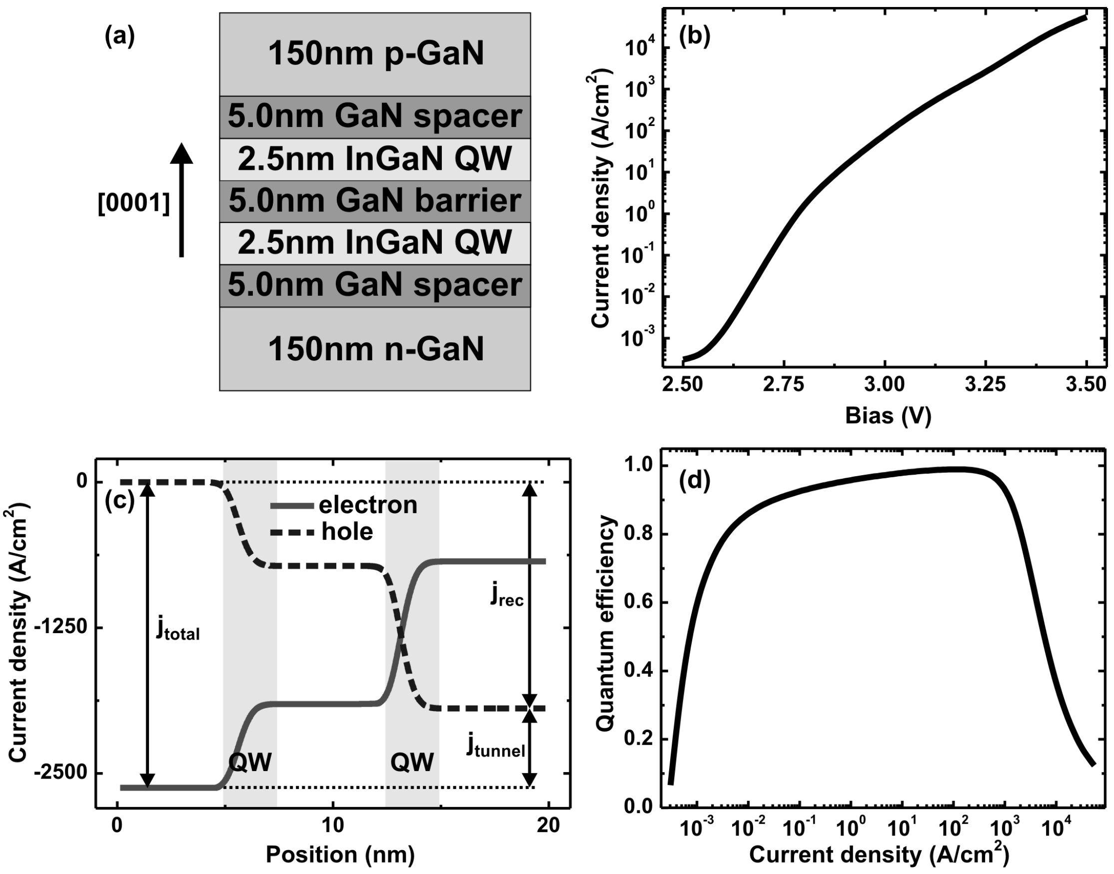


Figure 1.9: (a) Schematic cross section of $\mathrm{In}_{0.2}\mathrm{Ga}_{0.8}\mathrm{N}$ / GaN double quantum well LED. (b) Total current density j_{total} as a function of the applied bias. (c) Position dependent electron and hole current densities for a bias of 3.25 V. (d) Quantum efficiency ($j_{\mathrm{photo}}/j_{\mathrm{total}}$) as a function of j_{total}.

- At a "Fermi contact", we do not apply any boundary conditions on the Poisson equation. It is used to specify boundary conditions exclusively for the quasi Fermi levels, e.g., in an interior area of the device.

As an example, we apply the QDD method to a double well InGaN green LED that is sketched in Fig. 1.9(a). The device consists of two 2.5 nm $\mathrm{In}_{0.2}\mathrm{Ga}_{0.8}\mathrm{N}$ quantum wells that are separated by 5 nm GaN barriers. The layers are assumed to be grown on wurtzite [0001] n-GaN with a doping concentration of $N_D = 5 \cdot 10^{18}$ cm^{-3} and are finished by a 150 nm p-GaN ($N_A = 2 \cdot 10^{19}$ cm^{-3}) capping layer. We apply Ohmic contacts at the boundaries in the growth direction and calculate the current density as a function of the applied bias [see Fig. 1.9(b)]. Radiative- as well as Schockley-Read-Hall recombination processes have been considered in the calculation. We find the optical transition between the lowest electron and hole states in the quantum wells to lie at 420 nm in the green spectral range. In fact, not all of the current density

j_{total} leads to the creation of photons (j_{photo}). Besides non-radiative recombination losses ($j_{\text{non-rad}}$), for high bias values there is also a dark current (j_{tunnel}) of carriers that tunnel through the device without recombining,

$$j_{\text{total}} = j_{\text{rec}} + j_{\text{tunnel}} = j_{\text{photo}} + j_{\text{non-rad}} + j_{\text{tunnel}}. \tag{1.111}$$

This can be deduced from Fig. 1.9(c) which shows the position dependent current density for a bias of 3.25 V. As can be clearly seen, in each of the quantum wells a certain amount of electrons and holes recombine. However, a finite part remains that flows through the device and is therefore lost. This effect increases with the applied bias and limits the performance of the device. When drawing the quantum efficiency ($j_{\text{photo}}/j_{\text{total}}$) versus the total current density j_{total}, as is done in Fig. 1.9(d), one can extract the current range where the device works efficient. We find the present idealized LED to have a quantum efficiency of 99% for current densities up to 200 A/cm^2. Here, the photo-current has been determined by integrating the radiative recombination rate over the full length of the structure. Note that the inclusion of quantum mechanical charge densities is essential in the present structure, to obtain a finite tunnel-current through the barriers. So far, there exists no commercial software that can simulate such a LED due to their lack of quantum mechanics. With the original method from Ref. [5], the calculation also fails to converge. In next**nano**++, we have improved its numerical realization that will be discussed in Sec. 2.6.2 and now leads to rapid convergence.

1.6 Summary

In this chapter, we have presented the theoretical background for the calculation of realistic three-dimensional semiconductor nanostructures within next**nano**++. We have given an overview of the multiband **k·p** envelope function method (EFA) that is used for the calculation of the global electronic structure. The inclusion of magnetic fields has been briefly sketched, but will be discussed in detail in chapter 3. In addition, we have specified several inherent ambiguities and instabilities that come along with EFA and have shown solution strategies to eliminate these problems. We have expressed that doping-, free-carrier-, polarization- and fixed charges, result in a nonzero spatial distribution of excess charges. These charges are the source of a "built-in" electrostatic potential in a semiconductor nanostructure. Together with LSDA exchange correlation potentials and external electric fields, they are included in the Schrödinger equation via an external potential. Lattice mismatch between different semiconductor materials results in strain fields that are calculated in a continuum elasticity model and that modify the electronic structure in terms of deformation potentials and piezoelectric charges. Finally, an external electric field can be included by applying a bias between contacts. This drives the system out of thermodynamic equilibrium and causes electric currents to flow. Such a situation is described by position dependent quasi Fermi levels in a quantum drift-diffusion model.

Chapter 2

Numerical realization

2.1 Introduction

In this chapter, we present the numerical realization of the theoretical concepts introduced in the previous chapter. We start with an overview of the design of next**nano**++ [4] which is a successor of the next**nano**^{3} code [5]. This section basically describes the organizational part of the software as well as the code structure. Besides such logical arithmetics, a major part of the software is the numerical solution of physical relations. To this end, all partial differential equations are mapped on a discrete lattice by a discretization of the differential operators and therefore transformed into linear algebra problems. Afterwards, we describe algorithms that are used to solve the resulting systems of linear equations and matrix eigenvalue problems. Importantly, a reliable method for the solution of the eight-band Schrödinger equation is required, because this $\mathbf{k}\cdot\mathbf{p}$ model is frequently used throughout this thesis. Since in any realistic system, there are many connections between different physical aspects, we also face self-consistent problems. We therefore provide solution strategies for such coupled systems of equations.

2.2 Design of nextnano++

The program flow of next**nano**++ shown in Fig. 2.1 can be summarized as follows. The user specifies the input using a text file that defines the geometry and the materials of the nanostructure, the contact bias, and all other information needed to describe the physical system under consideration. This input file uses a concise hierarchical syntax that allows even very complex device geometries to be characterized by only a few lines. Then, in the initial phase, the program maps the geometry of the structure to a finite grid using a database that provides bulk parameters for a large number semiconductor materials. Next, next**nano**++ evaluates the bulk band structures of all constituent materials, performs a global strain calculation, and determines the new band edges and piezoelectric charges. Subsequently, the coupled

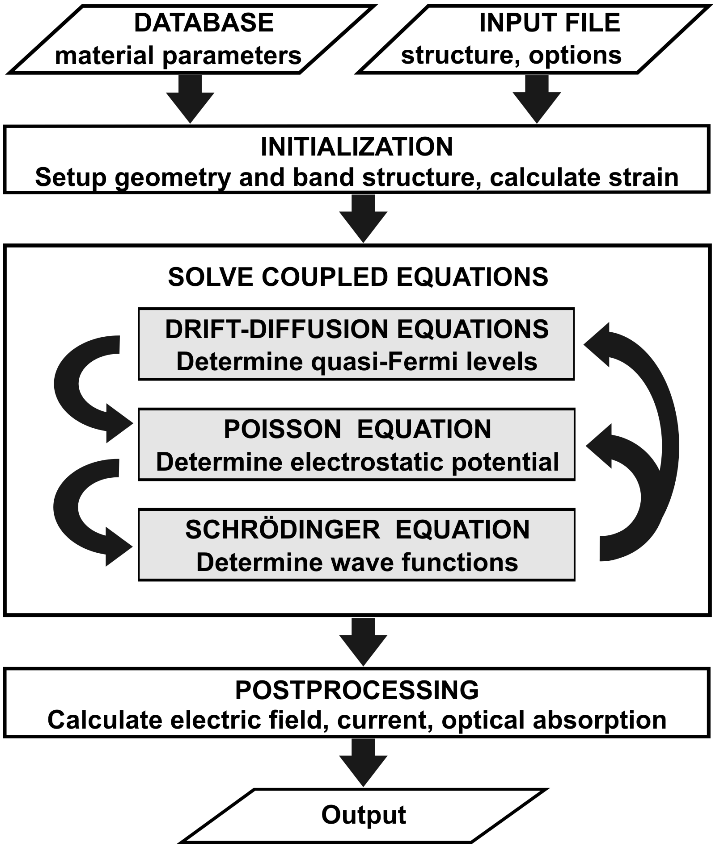


Figure 2.1: Flow scheme of next**nano**++.

Poisson-, current-, and multiband Schrödinger equations are solved self-consistently. This part is indicated by the box "Solve coupled equations" in Fig. 2.1 that shows the dependencies between the coupled systems of equations. Adequate solution strategies for this block will be detailed in Sec. 2.6. Finally, in a post-processing step, electric fields, drift-diffusion- as well as ballistic currents (CBR) [6], and optical matrix elements [47] can be computed. The output is generated in terms of ASCII or binary data files. In addition, we provide steering files that ease the visualization with commercial software tools such as Origin (1D and 2D data) or AVS/Express (2D and 3D data).

Originally, next**nano** was written in Fortran 90. But as the code (termed next**nano**^{3}) grew organically instead of in a planned fashion, it eventually consisted of 250,000 lines of very difficult to maintain code and a complete rewrite became necessary. Therefore, we have developed the new version next**nano**++ that is written mostly in C++ except for core numerical libraries and that uses modern object oriented programming techniques such as classes, inheritance, and generic programming in order to avoid code duplication. By this, the total amount of code has been suc-

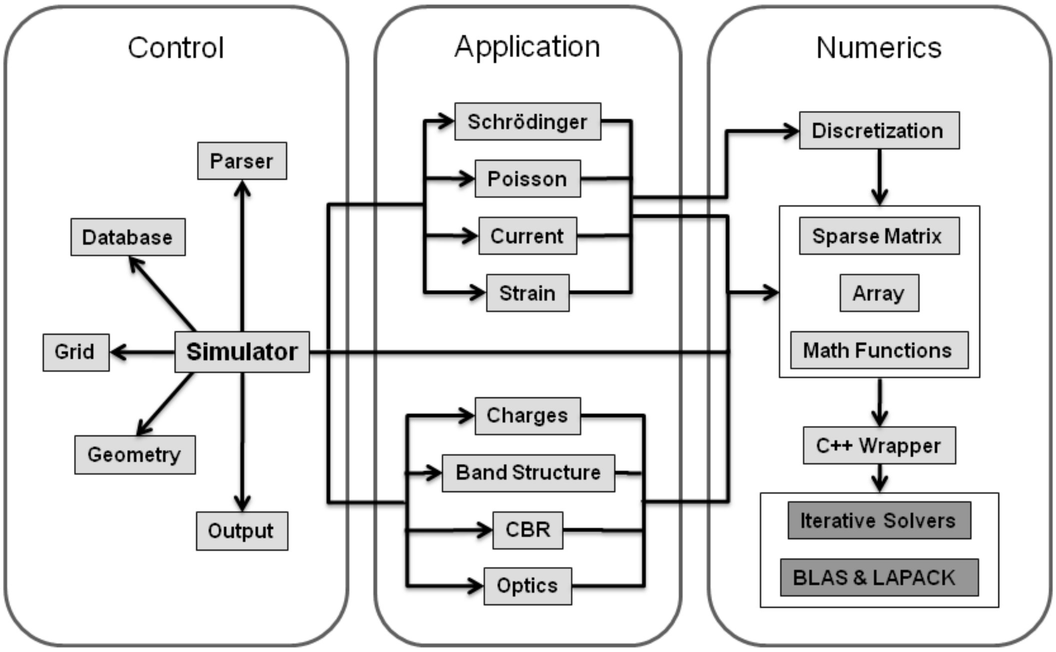
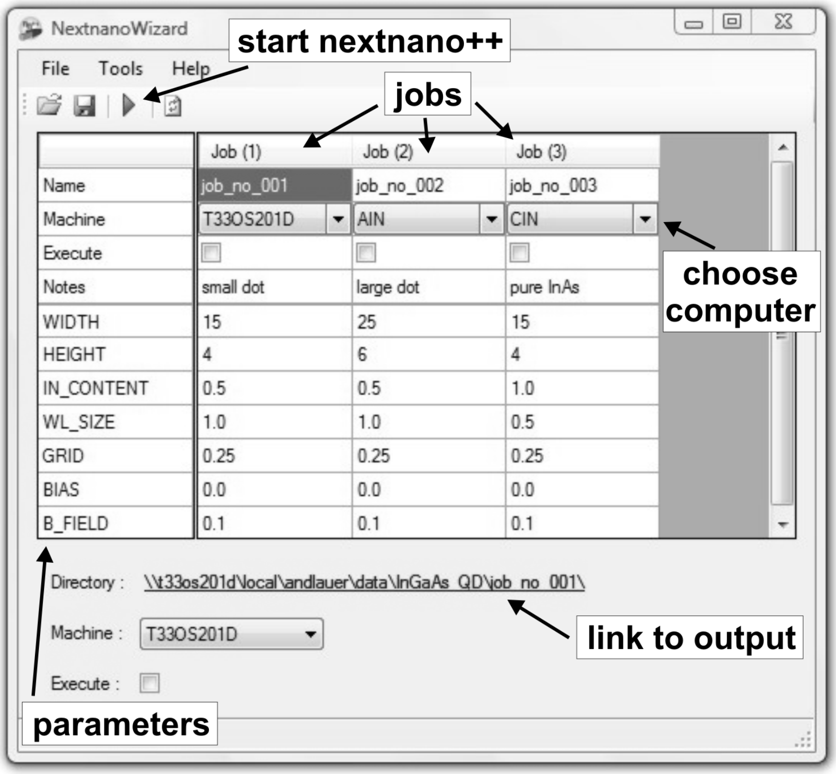


Figure 2.2: Overview of next**nano**++ code structure. The arrows indicate the dependencies between the individual modules. The software is written mostly in C++ (light gray) with only the core numerical libraries still being in Fortran (dark gray).

cessfully reduced by about a factor of five. In Fig. 2.2, we give an overview of the code structure used in next**nano**++. A more detailed description can be found in chapter 3 of Ref. [6]. The software is built in a modular way from multiple libraries that can be divided into control, application, and numeric ones. The base part of the software is the class "Simulator" that controls the main data- and program flow using the individual modules indicated by the boxes in the figure. These modules have partly been referred to already in the previous paragraph about the program flow. The input files are parsed using a Bison generated "Parser" module and validated for errors using an approach similar to the one used for validating XML files. The application group consists of libraries that implement the physical concepts presented in Chapter 1. All libraries that are based on partial differential equations ("Schrödinger", "Poisson", "Current", "Strain") access the same "Discretization" library that has been developed in a generalized, dimension independent way. Arrays and sparse matrices used within next**nano**++ are provided by appropriate libraries that implement multi-dimensional arrays and corresponding operations using a Fortran-compatible indexing style. These libraries as well as the "Discretization" are templated libraries, i.e., they use the same code for all possible data types such as integer, real, and complex numbers. Another example for efficient code reuse is the "Schrödinger" library that strongly relies on inheritance. Here, the data members and routines which all



	Job (1)	Job (2)	Job (3)
Name	job_no_001	job_no_002	job_no_003
Machine	T33OS201D	AIN	CIN
Execute	☐	☐	☐
Notes	small dot	large dot	pure InAs
WIDTH	15	25	15
HEIGHT	4	6	4
IN_CONTENT	0.5	0.5	1.0
WL_SIZE	1.0	1.0	0.5
GRID	0.25	0.25	0.25
BIAS	0.0	0.0	0.0
B_FIELD	0.1	0.1	0.1

Directory : \\t33os201d\local\andlauer\data\InGaAs_QD\job_no_001\

Machine : T33OS201D

Execute : ☐

Figure 2.3: Screenshot of NextnanoWizard.

different **k·p** models have in common, are introduced in a single base class. Now, each individual **k·p** model has a specialized Schrödinger class that inherits all general properties from the base class so that only the specific properties need to be implemented in the specialized class. Only the core numeric libraries such as iterative solvers, BLAS (Basic Linear Algebra Subprograms [48]) and LAPACK (Linear Algebra Package [49]) are still written in Fortran. They can be accessed within next**nano**++ via a templated C++ wrapper.

Recently, we have developed a convenient graphical user interface for next**nano**++ that is called NextnanoWizard. It needs a previously created template input file to be run with. This template is a generalized input file with a specified set of parameters that are meant to be modified. NextnanoWizard organizes these parameter sets like a conventional spreadsheet program, creates the appropriate input files and starts next**nano**++ which itself is a console application. Fig. 2.3 shows a screenshot of this novel software tool. The leftmost column in the figure lists the adjustable parameters, while the remaining columns represent concrete parameter sets that characterize specific devices and are termed "jobs". With the help of an additional software tool (termed VisRemote) developed by Peter Greck, the individual jobs can be executed remotely on designated computers in a network. Already completed jobs are stored in the template file for later reinvestigation. The output directories can be easily accessed via the link below the spreadsheet. In summary, NextanoWizard makes systematic sweeps over certain parameter spaces and book-keeping of previous results much easier to operate.

2.3 Discretization of differential operators

The numerical methods used in next**nano**++ are dictated by the simple fact that all equations we need to solve are partial differential equations (PDEs) in position space. Here, we employ box integration finite differences for the discretization [35, 50]. Compared to standard finite differences, this method has the advantage that discontinuities in material parameters (which often occur in semiconductor nanostructures), are naturally taken into account, since this method is flux conservative due to the Gauss theorem. In box discretization, the simulation domain is split into non-overlapping rectangular boxes, each covering the area closest to one grid node as indicated in Fig. 2.4. Any differential equation is now integrated over these box volumes to obtain an integral equation that is discretized using the finite differences scheme. The grid itself is always assumed to be a nonuniform tensor product grid that allows to concentrate nodes in regions of special interest. Such tensor grids have the advantage that no complicate meshing algorithm is required and the discretization can be easily and very efficiently implemented.

Since we are dealing with a large variety of differential equations that need to be solved for quasi one-, two-, or three-dimensional (in general d-dimensional) structures, a universal formulation for the discretization is highly desirable. To this end, we write any $n \times n$ differential matrix operator $\widehat{\mathcal{A}}\left(\mathbf{x}\right)$ of up to second-order in the following form,

$$\widehat{\mathcal{A}}\left(\mathbf{x}\right) = \sum_{i,j} \partial_i \widehat{\mathcal{A}}_{ij}\left(\mathbf{x}\right) \partial_j + \sum_{i} \left[\partial_i \widehat{\mathcal{A}}_{i\rightarrow}\left(\mathbf{x}\right) + \widehat{\mathcal{A}}_{i\leftarrow}\left(\mathbf{x}\right) \partial_i \right] + \widehat{\mathcal{A}}_0\left(\mathbf{x}\right). \tag{2.1}$$

Here, we have decomposed the operator $\widehat{\mathcal{A}}\left(\mathbf{x}\right)$ into $n \times n$ matrices $\widehat{\mathcal{A}}_{ij}\left(\mathbf{x}\right)$, $\widehat{\mathcal{A}}_{i\rightarrow}\left(\mathbf{x}\right)$, $\widehat{\mathcal{A}}_{i\leftarrow}\left(\mathbf{x}\right)$, and $\widehat{\mathcal{A}}_0\left(\mathbf{x}\right)$ (in general $\widehat{\mathcal{A}}_\gamma\left(\mathbf{x}\right)$ with $\gamma \in \{ij, i\rightarrow, i\leftarrow, 0\}$) that contain material-dependent parameters and are of different orders in the derivatives ∂_i (the indices $i = 1, ..., d$ denote the Cartesian components). The matrices $\widehat{\mathcal{A}}_\gamma\left(\mathbf{x}\right)$ uniquely define $\widehat{\mathcal{A}}\left(\mathbf{x}\right)$, so that we have effectively split the physical information from the topology which is given by the derivatives ∂_i. Now, we can discretize the derivatives ∂_i and take the matrices $\widehat{\mathcal{A}}_\gamma\left(\mathbf{x}\right)$ as parameters. This leads to a discretization scheme that becomes independent from the concrete form of $\widehat{\mathcal{A}}\left(\mathbf{x}\right)$.

In order to apply the box discretization, we introduce the following concise notation. We consider a nonuniform rectangular grid that is oriented parallel to the Cartesian axes. We assume the space to be d-dimensional generally and enumerate every grid point by a d-dimensional tuple $\mathbf{m}$ of integers. Now, the position in space of

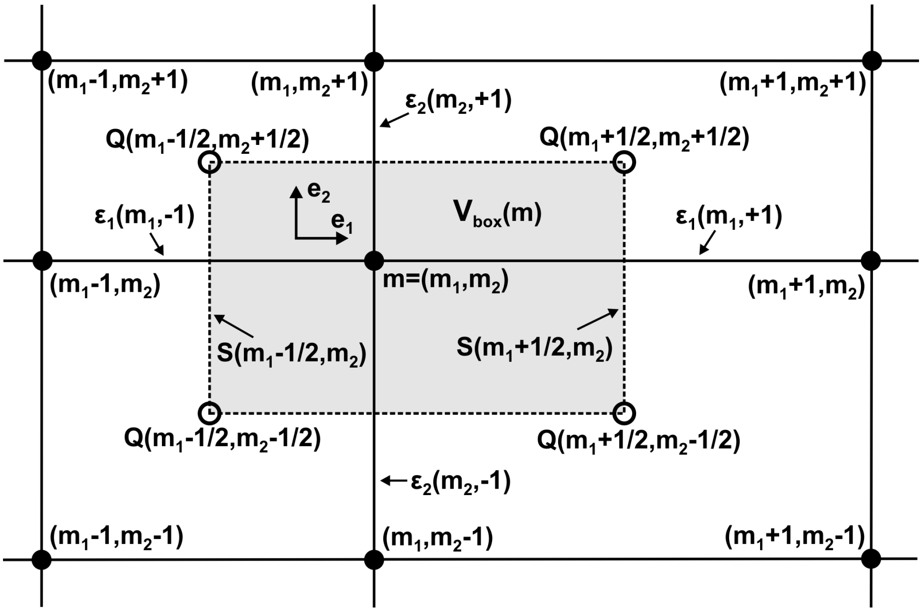


Figure 2.4: Sketch of two-dimensional inhomogenous rectangular grid. For box discretization, each grid node $\mathbf{m}$ is surrounded by a control box $V_{\text{box}}(\mathbf{m})$ (gray region). Each lattice area $Q(m_1 \pm 1/2, m_2 \pm 1/2)$ that is formed by four nearest-neighbor grid nodes is assumed to be filled by a constant material.

the point $\mathbf{m}$ is given by the d-dimensional vector $\mathbf{x}(\mathbf{m})$. For simplicity, we only take into account nearest-neighbor interactions. So we need to address neighbor points $\mathbf{n}$ that deviate from $\mathbf{m}$ in each coordinate by a maximum of one. To this end, we introduce the notation

$$\mathbf{x}(\mathbf{n}) = \mathbf{x}(\mathbf{m}, \mathbf{s}) \equiv \mathbf{x}(m_1 + s_1, ..., m_d + s_d), \tag{2.2}$$

where the vector $\mathbf{s}$ consists of the elements $s_i \in \{0, \pm 1\}$. In order to express that a certain point $\mathbf{n}$ differs from $\mathbf{m}$ explicitly in one coordinate i or two coordinates i, j, we may write $\mathbf{n} = (\mathbf{m}, s'_i)$ or $\mathbf{n} = \left(\mathbf{m}, s'_i, s'_j\right)$, respectively. Here the "dash" means that the zero is excluded so that $s'_i, s'_j \in \{\pm 1\}$. We further define grid spacings between nearest-neighbor points

$$\varepsilon_i(m_i, s'_i) = s'_i \left[x_i(\mathbf{m}, s'_i) - x_i(\mathbf{m})\right]. \tag{2.3}$$

Note that this grid spacing does not depend on the indices $m_{j \neq i}$ since the grid is rectangular. The grid lines split the simulation domain into rectangular lattice areas $Q(\mathbf{m}, \mathbf{s}'/2)$ that may be labeled by their center point using intermediate-indices, as has been depicted in Fig. 2.4. We assume each of these areas to be filled by a homogenous material so that any material parameter $a(\mathbf{x})$ is constant for each $\mathbf{x} \in Q(\mathbf{m}, \mathbf{s}'/2)$. By this, our grid is placed on the simulation domain in a way that all material interfaces coincide with grid lines.

In the process of discretization, any differential operator $\partial_i f(\mathbf{x})$ is replaced by a discrete approximation $\delta_i f(\mathbf{m})$. We will now illustrate the box discretization approach by means of a second-order operator

$$\partial_1 a(\mathbf{x})\,\partial_1 f(\mathbf{x}) \to \delta_1 a(\mathbf{m})\,\delta_1 f(\mathbf{m}), \tag{2.4}$$

for the two-dimensional grid shown in Fig. 2.4. As can be seen from the figure, for each grid node $\mathbf{m} = (m_1, m_2)$ we introduce a two-dimensional rectangular area

$$V_{\text{box}}(\mathbf{m}) = \frac{\varepsilon_1(m_1, +1) + \varepsilon_1(m_1, -1)}{2} \times \frac{\varepsilon_2(m_2, +1) + \varepsilon_2(m_2, -1)}{2}, \tag{2.5}$$

which will be termed the control box (we use the term V for a general volume, although it is in fact an area in the two-dimensional case). For the discretization, we first integrate the expression in Eq. (2.4) over the control box

$$\partial_1 a(\mathbf{x})\,\partial_1 f(\mathbf{x}) \to \frac{1}{V_{\text{box}}(\mathbf{m})} \int_{V_{\text{box}}(\mathbf{m})} d^2\mathbf{x}\,\partial_1 a(\mathbf{x})\,\partial_1 f(\mathbf{x}), \tag{2.6}$$

for each grid node. Using Gauss's theorem, this surface integral can be cast into two line integrals (in three-dimensions, a volume integral is cast into surface integrals)

$$\frac{1}{V_{\text{box}}(\mathbf{m})} \int_{V_{\text{box}}(\mathbf{m})} d^2\mathbf{x}\,\partial_1 a(\mathbf{x})\,\partial_1 f(\mathbf{x}) = \frac{1}{V_{\text{box}}(\mathbf{m})} \sum_{s_1'=\pm1} s_1' \int_{S\left(\mathbf{m}, s_1'/2\right)} dx_2\, a(\mathbf{x})\,\partial_1 f(\mathbf{x}), \tag{2.7}$$

where the lines $S(m_1 \pm 1/2, m_2)$ denote the boundary segments of the control box in the positive and negative Cartesian $\mathbf{e}_1$-direction, respectively. We now use finite differences to approximate the derivatives at the center of each boundary segment

$$\delta_1 f(m_1 \pm 1/2, m_2) = \pm\frac{f(m_1 \pm 1, m_2) - f(m_1, m_2)}{\varepsilon_1(m_1, \pm1)}. \tag{2.8}$$

Taking into account that $a(\mathbf{x})$ is constant within each of the four quadrants $Q(m_1 \pm 1/2, m_2 \pm 1/2)$ of the control box, we can easily evaluate the remaining integrals

$$\int_{S(m_1\pm1/2, m_2)} dS\, a(\mathbf{x}) = \sum_{s_2'=\pm1} \frac{1}{2}\varepsilon_2(m_2, s_2')\, a(m_1 \pm 1/2, m_2 + s_2'/2). \tag{2.9}$$

Altogether, we obtain the following discretization formula

$$\delta_1 a(\mathbf{m})\,\delta_1 f(\mathbf{m}) = \frac{1}{V_{\text{box}}(\mathbf{m})} \sum_{s_1'=\pm1, s_2'=\pm1} a(\mathbf{m}, \mathbf{s}'/2)\,\frac{\varepsilon_2(m_2, s_2')}{2\varepsilon_1(m_1, s_1')}\left[f(\mathbf{m}, s_1') - f(\mathbf{m})\right]. \tag{2.10}$$

This expression can also be generalized for an arbitrary direction i and for a d-dimensional grid. The derivation has been performed in every detail in Ref. [28] and results in

$$\delta_i a(\mathbf{m})\,\delta_i f(\mathbf{m}) = \frac{1}{V_{\text{box}}(\mathbf{m})}\sum_{\mathbf{s}'} a(\mathbf{m},\mathbf{s}'/2)\,\frac{\prod_{j\neq i}\varepsilon_j\left(m_j,s'_j\right)}{2^{d-1}\varepsilon_i\left(m_i,s'_i\right)}\left[f\left(\mathbf{m},s'_i\right)-f(\mathbf{m})\right], \tag{2.11}$$

with the d-dimensional box volume

$$V_{\text{box}}(\mathbf{m}) = \prod_i \frac{\sum_{s'_i}\varepsilon_i\left(m_i,s'_i\right)}{2}. \tag{2.12}$$

Here, the sum runs over all possible vectors $\mathbf{s}'$, that can be constructed when for each element s'_i two values ± 1 are allowed. By this, we add up the contributions of all 2^d adjacent material areas $Q(\mathbf{m},\mathbf{s}'/2)$. Similar formulas can be derived for all other operators in Eq. (2.1) (cf. Ref. [28]). In summary, these are given by

$$\begin{aligned}\delta_i a(\mathbf{m})\,\delta_j f(\mathbf{m}) = {} & \frac{1}{V_{\text{box}}(\mathbf{m})}\sum_{\mathbf{s}'} s'_i s'_j a(\mathbf{m},\mathbf{s}'/2)\,\frac{\prod_{k\neq i,j}\varepsilon_k\left(m_k,s'_k\right)}{2^d}\\ & \times\left[f\left(\mathbf{m},s'_i,s'_j\right)-f\left(\mathbf{m},s'_i\right)+f\left(\mathbf{m},s'_j\right)-f(\mathbf{m})\right],\ (j\neq i)\end{aligned} \tag{2.13}$$

$$a(\mathbf{m})\,\delta_i f(\mathbf{m})\big|_{\text{cent}} = \frac{1}{V_{\text{box}}(\mathbf{m})}\sum_{\mathbf{s}'} s'_i a(\mathbf{m},\mathbf{s}'/2)\,\frac{\prod_{j\neq i}\varepsilon_j\left(m_j,s'_j\right)}{2^d}\left[f\left(\mathbf{m},s'_i\right)-f(\mathbf{m})\right], \tag{2.14}$$

$$\delta_i a(\mathbf{m})\,f(\mathbf{m})\big|_{\text{cent}} = \frac{1}{V_{\text{box}}(\mathbf{m})}\sum_{\mathbf{s}'} s'_i a(\mathbf{m},\mathbf{s}'/2)\,\frac{\prod_{j\neq i}\varepsilon_j\left(m_j,s'_j\right)}{2^d}\left[f\left(\mathbf{m},s'_i\right)+f(\mathbf{m})\right], \tag{2.15}$$

$$\begin{aligned}a(\mathbf{m})\,\delta_i f(\mathbf{m})\big|_{\text{forw/backw}} = {} & \frac{1}{V_{\text{box}}(\mathbf{m})}\sum_{\mathbf{s}'} a(\mathbf{m},\mathbf{s}'/2)\,\frac{\prod_{j\neq i}\varepsilon_j\left(m_j,s'_j\right)}{2^d}\\ & \times\left[\left(s'_i\pm 1\right)f\left(\mathbf{m},s'_i\right)-\left(s'_i\pm 1\right)f(\mathbf{m})\right],\end{aligned} \tag{2.16}$$

$$\begin{aligned}\delta_i a(\mathbf{m})\,f(\mathbf{m})\big|_{\text{forw/backw}} = {} & \frac{1}{V_{\text{box}}(\mathbf{m})}\sum_{\mathbf{s}'} a(\mathbf{m},\mathbf{s}'/2)\,\frac{\prod_{j\neq i}\varepsilon_j\left(m_j,s'_j\right)}{2^d}\\ & \times\left[\left(s'_i\pm 1\right)f\left(\mathbf{m},s'_i\right)-\left(-s'_i\pm 1\right)f(\mathbf{m})\right],\end{aligned} \tag{2.17}$$

$$a(\mathbf{m})\,f(\mathbf{m}) = \frac{1}{V_{\text{box}}(\mathbf{m})}\sum_{\mathbf{s}'} a(\mathbf{m},\mathbf{s}'/2)\,\frac{\prod_i\varepsilon_i\left(m_i,s'_i\right)}{2^d}f(\mathbf{m}). \tag{2.18}$$

In fact, the discretization formulas have been derived in a dimension-independent way, where d basically reduces to a summation limit. This allows us to use the same discretization routines for all types of equations and spatial dimensions so that consistency is always guaranteed.

After discretization, all differential operators $\widehat{\mathcal{A}}(\mathbf{x})$ in the PDEs have now become $N \times N$ matrices $\hat{A}$, where N denotes the number of grid nodes, and any function $f(x)$ has become a N-vector $\mathbf{f}$. If $\widehat{\mathcal{A}}(\mathbf{x})$ is in fact a $n \times n$ matrix operator, $\hat{A}$ will have a $n \times n$ block structure with each block element being a $N \times N$ matrix. Since we include only nearest-neighbor interactions, $\hat{A}$ becomes a sparse matrix which due to the underlying tensor grid has a banded sparsity structure. Since we consider operators of up to second order in the derivatives ∂_i, the matrix elements $\hat{A}(\mathbf{m}, \mathbf{n})$ can be nonzero both for the $2d$ nearest as well as for the $2d(d-1)$ diagonally adjacent (next-nearest) neighbors. Now, every PDE that is a boundary value problem (such as the Poisson, the drift-diffusion-current, and the strain equation (1.65), (1.105), and (1.83), respectively) becomes a large system of linear (or non-linear) equations after discretization,

$$\hat{A}\mathbf{f} = \mathbf{b}, \tag{2.19}$$

while the $\mathbf{k}\cdot\mathbf{p}$ Schrödinger equation [Eq. (1.41)] yields a large matrix eigenvalue problem,

$$\hat{A}\mathbf{u} = \lambda\mathbf{u}, \tag{2.20}$$

with discretized wave functions $\mathbf{u}$. In the following sections, we will describe the methods that we have employed in next**nano**++ in order to solve these problems. Eventually, we are interested not only in the solutions $\mathbf{f}$ of Eq. (2.19), but also the gradients $\boldsymbol{\nabla}\mathbf{f}$ can form important observables. In Appendix D, the calculation of these properties will be detailed. Finally, we would like to note that for nonuniform grids, Eq. (2.20) in fact becomes a generalized eigenvalue problem

$$\hat{D}\hat{A}'\mathbf{u} = \lambda\mathbf{u}, \tag{2.21}$$

with $\hat{A} = \hat{D}\hat{A}'$, where $\hat{A}'$ is Hermitian and $\hat{D}$ is a diagonal matrix whose elements are determined by V_{box} [Eq. (2.12)]. However, applying the transformation

$$\hat{D}^{1/2}\hat{A}'\hat{D}^{1/2}\mathbf{u}' = \lambda\mathbf{u}', \tag{2.22}$$

with $\mathbf{u}' = \hat{D}^{-1/2}\mathbf{u}$ leads one back to a Hermitian eigenvalue problem.

2.4 Solution of linear and nonlinear systems of equations

When modeling complex three-dimensional structures, the number of required grid nodes can easily exceed $N = 100 \times 100 \times 100 = 10^6$. This makes it inevitable to use storage schemes that fully exploit the sparsity structure of the matrices in order to manage the large memory consumption. Storing only the nonzero elements of a matrix excludes Gaussian elimination from being a useful solution method for

Eq. (2.19), because this would fill up the matrix and requires $O\left(N^3\right)$ operations. Iterative methods that determine a sufficiently accurate approximate solution by a sequence of matrix vector multiplications (MATMULs) and little additional work can be applied much more effectively. Such methods obtain an additional profit from the sparsity structure that makes a MATMUL an $O\left(N\right)$ operation, contrary to general dense matrices where it is an $O\left(N^2\right)$ operation. Detailed information on a large number of iterative methods for solving systems of equations can be found in Ref. [51]. We will concentrate only on those methods that are actually used within next**nano**++.

2.4.1 Linear systems of equations

For Hermitian positive definite problems, we use the well known conjugate gradient method (CG) [52]. This method derives its name from the fact that it generates a sequence of *conjugate* (or orthogonal) vectors. These vectors are the *gradients* $\boldsymbol{\nabla} F$ of a quadratic functional

$$F = \frac{1}{2}\mathbf{f}^T\hat{A}\mathbf{f} - \mathbf{b}\mathbf{f}. \tag{2.23}$$

The Algorithm minimizes this functional,

$$\boldsymbol{\nabla} F = \hat{A}\mathbf{f} - \mathbf{b} \equiv 0, \tag{2.24}$$

which is equivalent to solving the linear system [Eq. (2.19)]. CG is our method of choice for the linear Poisson equation [Eq. (1.65)], the drift-diffusion current equation [Eq. (1.105)], and the strain equation [Eq. (1.83)] with non-periodic boundary conditions. It turns out that it often also works for indefinite problems such as the shift-invert spectral transformation that will be discussed in Sec. 2.5.2.

If CG fails, we fall back to a composite step conjugate gradient method (CSCG). This method has been developed specifically to deal with indefinite problems as well. It differs from CG only in its strategy to combine certain iterations to *composite steps*, in order to improve the stability of the convergence process [53]. This method is not our standard method since it is significantly slower than CG.

For non-symmetric matrices, such as the strain equation with periodic boundary conditions, we use a biconjugate gradient method (BiCG) [54]. This method generates two CG-like sequences of vectors, one based on a system with the original coefficient matrix $\hat{A}$, and one on $\hat{A}^T$. Instead of orthogonalizing each sequence separately, they are made mutually orthogonal, or “bi-conjugate”. BiCG is much slower than CG, because it requires a multiplication with $\hat{A}$ and $\hat{A}^T$ in each iteration.

In order to improve convergence, we do not use the methods stated above in their original form, but accelerate them by preconditioners. All of those iterative methods converge very fast if the matrix $\hat{A}$ is close to the identity. Unfortunately, in most applications this is not the case. Therefore we replace the original linear system Eq. (2.19) by the modified system

$$\hat{M}^{-1}\hat{A}\mathbf{f} = \hat{M}^{-1}\mathbf{b}, \tag{2.25}$$

where the preconditioner matrix $\hat{M}$ is chosen so that $\hat{M}^{-1}\hat{A}$ approximates the identity. The speed of the whole algorithm is improved, when the benefit from the (easier to solve) modified problem is larger that the extra cost for constructing and applying the preconditioner. When designing a preconditioner, one is faced with a choice between finding a matrix $\hat{M}$ that approximates $\hat{A}$ and for which solving a system is easier than solving one with $\hat{A}$, or finding a matrix $\hat{M}$ that approximates $\hat{A}^{-1}$ so that only a multiplication by $\hat{M}$ is needed. In next**nano**++, we use preconditioners from the first category, because approximations of $\hat{A}^{-1}$ require storage that grows at least with $O\left(N^{4/3}\right)$, making them unsuitable for very large problems. In fact, we usually employ an incomplete Cholesky factorization [55], or for Poisson operators, the very fast fill-in-free Dupont-Kendall-Rachford method [56]. Both methods are based on incomplete factorizations of the coefficient matrix. Such a preconditioner is then given in factored form $\hat{M} = \hat{L}\hat{U}$ with a lower triangular matrix $\hat{L}$ and an upper triangular matrix $\hat{U}$. In the symmetric case it can be written $\hat{M} = \hat{L}\hat{D}\hat{L}^T$ with a diagonal matrix $\hat{D}$. The factorization is called incomplete, if during the factorization process certain fill elements (nonzero elements in positions where the original matrix had a zero) have been ignored.

2.4.2 Nonlinear systems of equations

If the original system [Eq. (2.19)] is nonlinear in the solution vector,

$$\hat{A}\mathbf{f} = \mathbf{b}\left(\mathbf{f}\right), \tag{2.26}$$

the methods presented above are not applicable. In next**nano**++, this situation arises for the Poisson equation [Eq. (1.65)]

$$\boldsymbol{\nabla}\varepsilon\boldsymbol{\nabla}\phi = -\rho\left(\phi\right), \tag{2.27}$$

where the non-linearity is introduced by the dependence of the classical charge density $\rho = \rho_{\mathrm{cl}}$ [Eq. (1.61)] on the electrostatic potential ϕ. In order to solve Eq. (2.26), we use the Newton-Raphson method with inexact line search. Here, we rewrite the problem into finding the root of a function

$$\mathbf{F}\left(\mathbf{f}\right) \equiv \hat{A}\mathbf{f} - \mathbf{b}\left(\mathbf{f}\right) = 0. \tag{2.28}$$

Starting from an initial guess $\mathbf{f}_0$, the solution is approached following the iteration sequence

$$\mathbf{f}_{k+1} = \mathbf{f}_k - \lambda\mathbf{g}_k, \tag{2.29}$$

where $\mathbf{g}_k$ denotes the direction of steepest descent of $\mathbf{F}$ at the position $\mathbf{f}_k$ which we follow for a step length λ. By simple algebra one can derive that $\mathbf{g}_k$ is given by

$$\mathbf{g}_k = \left[\hat{J}_{\mathbf{F}}\left(\mathbf{f}_k\right)\right]^{-1}\mathbf{F}\left(\mathbf{f}_k\right), \tag{2.30}$$

with the Jacobian matrix $\hat{J}_{\mathbf{F}}(\mathbf{f}) = \hat{A} - (D_{\mathbf{f}}\mathbf{b})(\mathbf{f})$. Rather than actually computing the inverse of this matrix, we solve the system of equations in each iteration

$$\hat{J}_{\mathbf{F}}(\mathbf{f}_k) \cdot \mathbf{g}_k = \mathbf{F}(\mathbf{f}_k)\,. \tag{2.31}$$

So effectively, this algorithm remaps the nonlinear problem into a sequence of linear solutions that can be computed the way it has been presented in Sec. 2.4.1. The remaining problem is to find a suitable step length λ for Eq. (2.29). Since plain Newton-Raphson iteration ($\lambda = 1$) may fail if the initial guess is not sufficiently close to the final solution, we stabilize the convergence by a line search [34]. In order to limit the computation cost, we employ this line search only in an incomplete manner. So we first try one full Newton step in the direction of $\mathbf{g}_k$ and keep dividing ($\lambda = 1/2, 1/4, ...$) or multiplying ($\lambda = 2, 4, ...$) the step size by 2 until a local minimum in the residual $\|\mathbf{F}(\mathbf{f}_{k+1})\|$ has been found.

2.5 Solution of matrix eigenvalue problems

Solving the Schrödinger equation for realistic nanostructures requires the solution of very large matrix eigenvalue problems [Eq. (2.20)], which is computationally very demanding. Here, memory consumption and CPU time pose strong limitations on the number of eigenvectors that can be calculated. Fortunately, in most situations only a few eigenstates are physically relevant, because the occupations of electron and hole states fall off exponentially with the energy distance from the Fermi level. This allows us to rely on iterative eigensolvers that utilize the so called power method and can be accelerated by spectral transformation preconditioners. A detailed description of the most common solution algorithms can be found in Ref. [57].

There are two types of eigenvalue problems that have to be solved in next**nano**++, namely extremal and non-extremal ones. Extremal eigenvalue problems come with the single-band and the six-band Schrödinger equations. Here, only the eigenvalues that lie closest to one end of the energy spectrum need to be calculated. This type of eigenvalue problem can be solved using an Arnoldi iteration method that has proven to be fast and robust. The eight-band Schrödinger equation, however, couples electron and hole bands and therefore leads to a non-extremal eigenvalue problem. Since the Fermi level lies in the vicinity of the energy gap, we are usually interested in the eigenstates close to this energy gap that itself lies somewhere in the middle of the energy spectrum of the Hamiltonian matrix. This limits the range of possible solution algorithms.

Since developing a fast and robust eigenvalue package is a very challenging and time consuming task, we rely on already available software libraries and customize these for our needs. In next**nano**++, we use three types of eigenvalue solvers; namely an implicitly restarted Arnoldi iteration solver from the ARPACK library [58], a Jacobi-Davidson iteration solver [59] and a direct dense matrix solver from the LAPACK library [49].

2.5.1 Extremal eigenvalue problems

For extremal eigenvalue problems, we rely on the ARPACK solver [58]. It is an iterative method that utilizes the power method which is based on the following principle. Since the matrix $\hat{A}$ in Eq. (2.20) is Hermitian, it can be written in terms of its eigenvectors $\mathbf{u}_i$ and corresponding eigenvalues λ_i,

$$\hat{A} = \sum_i \lambda_i \mathbf{u}_i \mathbf{u}_i^T. \tag{2.32}$$

Furthermore, any normalized trial vector $\mathbf{v}_0$ can be expanded into the eigenvectors of $\hat{A}$,

$$\mathbf{v}_0 = \sum_j c_j \mathbf{u}_j. \tag{2.33}$$

Then the multiplication $\hat{A}\mathbf{v}_0$ yields

$$\mathbf{v}_1 \equiv \hat{A}\mathbf{v}_0 = \sum_j \lambda_j c_j \mathbf{u}_j. \tag{2.34}$$

Hereby, the component of the largest eigenvalue $\lambda_{\max} \in \{\lambda_i\}$ is amplified most, so that many successive multiplications k (with subsequent normalization) will convert $\mathbf{v}_0$ into $\mathbf{u}_{\max}$,

$$\mathbf{v}_k \equiv \hat{A}^k\mathbf{v}_0 = \sum_j \lambda_j^k c_j \mathbf{u}_j \approx \lambda_{\max}^k c_{\max} \mathbf{u}_{\max}. \tag{2.35}$$

Now, the next eigenvector could be obtained by removing $\lambda_{\max}$ from the spectrum

$$\hat{A} \rightarrow \hat{A}' = \hat{A} - \lambda_{\max} \mathbf{u}_{\max} \mathbf{u}_{\max}^T, \tag{2.36}$$

and subsequently restarting the algorithm. However, it is more effective to store the individual steps in the so called Krylov subspace to find multiple eigenvalues simultaneously. In fact, ARPACK does this by using Arnoldi iterations in the Krylov subspace $\mathcal{K}_n$. Starting with the initial normalized guess vector $\mathbf{v}_0$, first an orthonormal basis of

$$\mathcal{K}_n = \operatorname{span}\left\{\mathbf{v}_0, \hat{A}\mathbf{v}_0, \hat{A}^2\mathbf{v}_0, ..., \hat{A}^n\mathbf{v}_0\right\}, \tag{2.37}$$

is calculated. Afterwards, the eigenvalues of the orthogonal projection of $\hat{A}$ onto the basis of $\mathcal{K}_n$ are computed. These so called Ritz eigenvalues converge to the extremal eigenvalues of $\hat{A}$. In order to limit the storage cost required for maintaining orthogonality of basis vectors with increasing number of iterations, the Arnoldi method is restarted after a few iterations. Therefore, the size n of the Krylov subspace needs to be chosen only two or three times larger than the number of desired eigenvectors in order to achieve a sufficiently precise solution.

For the huge matrices that occur in three-dimensional calculations, however, additional speed up of ARPACK is highly desirable and can be achieved by an additional

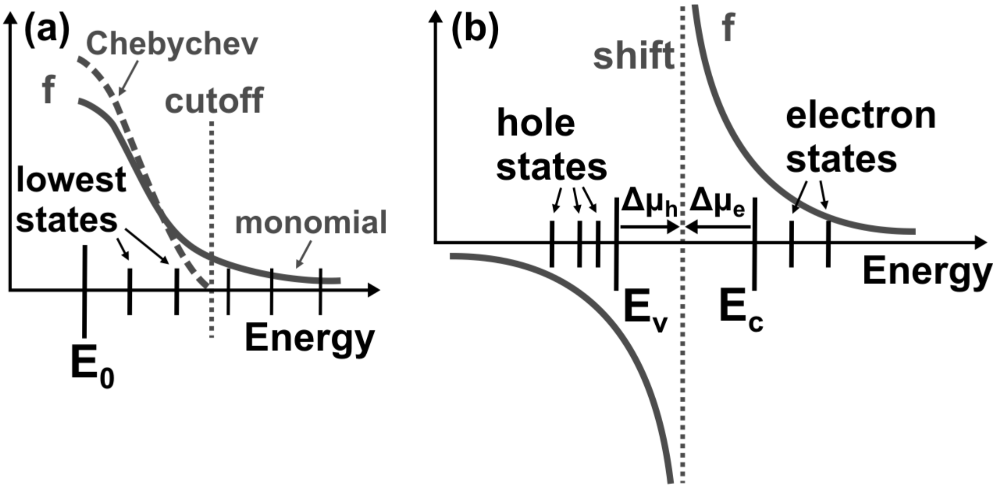


Figure 2.5: Sketches of spectral transformation functions f for (a) monomial (solid line) and Chebychev polynomial (dashed line) transformation, (b) shift-invert transformation.

preconditioner. Here, the aim of preconditioning is to spread the spectrum of $\hat{A}$ in the interesting range of eigenvalues and to compress it elsewhere. Such a preconditioner can be obtained utilizing a polynomial function f for a spectral transformation,

$$\hat{A}\mathbf{u} = \lambda\mathbf{u} \Longleftrightarrow f\left(\hat{A}\right)\mathbf{u} = f\left(\lambda\right)\mathbf{u}. \tag{2.38}$$

The basic idea behind such a spectral transformation is the fact that the eigenvectors $\mathbf{u}$ of a matrix $\hat{A}$ are invariant under this transformation, whereas the eigenvalues λ do change and can be amplified at the end of the spectrum where the polynomials are chosen to be large (see Fig. 2.5). This improves conditioning and thus accelerates the convergence of these extremal eigenvalues. In next**nano**++, we use simple monomials x^n, Chebychev polynomials, and Legendre polynomials. In most situations, we find the Chebychev polynomials to provide the best speed-up that can reach a factor of ten compared to the standard Arnoldi method. The Chebychev polynomials outperform the simple monomials by about a factor of two, because they can amplify the extremal part of the spectrum more effectively. However, they require to define an energy cutoff above which no eigenvalues can be computed any more. This has been sketched in Fig. 2.5(a). It is difficult to predict a reasonable energy cutoff in advance, because a value that lies too close to the end of the spectrum will prevent convergence, while a value that lies too far from the relevant part of the spectrum annihilates the performance improvement of the method. Therefore, we use the following robust strategy for charge-self-consistent problems. When the Schrödinger equation is solved for the first time, we always apply the monomial preconditioner that does not require any knowledge of the relevant energy range of the spectrum. In all further steps that require solving the Schrödinger equation, we use the Chebychev preconditioner with an energy cutoff that lies slightly above the highest calculated eigenvalue from the last iteration (a value of 0.1 eV has proven to be sufficient in most situations). If ARPACK fails nonetheless using this cutoff, we fall back to the monomial preconditioner to guarantee convergence in any case.

2.5.2 Non-extremal eigenvalue problems

For non-extremal eigenvalue problems that come with the eight-band Schrödinger equation, the method described above it not directly applicable and other solution algorithms are needed. For small one-dimensional problems, we can rely on a direct dense matrix solver from the LAPACK library [49] that has proven to be fast and robust. However, with increasing problem size (quasi two- or three-dimensional structures) this method quickly becomes ineffective, because its dense storage scheme and its direct diagonalization algorithm obviously scale very poorly with the number of grid nodes. For very large matrix eigenvalue problems, iterative methods can easily outperform direct ones by orders of magnitude. In this case, a Jacobi-Davidson iteration method [59] can be applied successfully. We find this method to be reasonably fast, but rather difficult to control. It requires an energy shift to be defined that lies close to the relevant matrix eigenvalues and convergence speed drops down drastically with increasing difference between this energy guess and the true eigenvalues. It is difficult to predict a good energy guess in advance, especially if the relevant eigenvalues comprise a larger energy interval. In the latter case, the algorithm has to be restarted several times with different guess energies. Together with careful book-keeping of already converged eigenvalues, this is rather complicated to automatize. By modifying the ARPACK solver with a shift-invert spectral transformation (ARPACK-SI), we have successfully managed to apply the Arnoldi iteration method to non-extremal eigenvalue problems as well. We find this algorithm to be faster and more reliable than the Jacobi-Davidson method in our applications and will therefore detail its concept in the following. In next**nano**++, we fall back to the Jacobi-Davidson method only if the ARPACK-SI algorithm fails.

The shift-invert spectral transformation is given by

$$\hat{A}\mathbf{u} = \lambda\mathbf{u} \Longleftrightarrow \left(\hat{A} - \mu\hat{I}\right)^{-1}\mathbf{u} = (\lambda - \mu)^{-1}\mathbf{u}, \tag{2.39}$$

where μ denotes an energy shift. This corresponds to a spectral transformation function that has a singularity at the position of the shift, as shown in Fig. 2.5(b). Therefore, the inner eigenvalues λ close to μ become extremal eigenvalues lying either at the lower end (eigenvalues slightly below μ become strongly negative) or at the upper end of the spectrum. By this transformation, ARPACK can be used to calculate the non-extremal eigenvalues in the vicinity of μ. Since inverting a large sparse matrix would create a dense matrix that exceeds any realistically available amount of memory, we have to avoid this operation by solving a system of linear equations,

$$\left(\hat{A} - \mu\hat{I}\right)\mathbf{u}_{k+1} = \mathbf{u}_k, \tag{2.40}$$

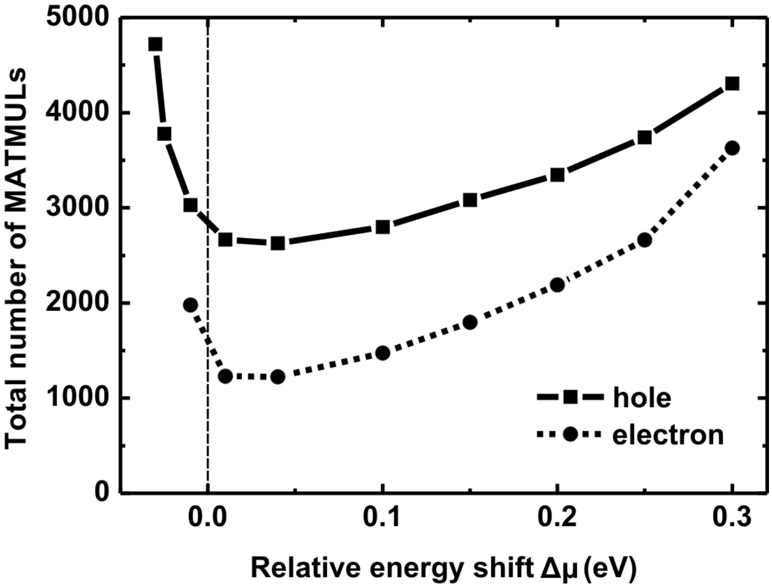


Figure 2.6: Total number of matrix-vector multiplications that are required for the calculation of the eight-band hole (solid line) and electron (dotted line) ground states of an InAs pyramidal quantum dot, as a function of the energy shift.

in each Arnoldi iteration. This is more time consuming than the polynomial transformations for extremal eigenvalue problems. Although, the matrix in Eq. (2.40) is indefinite, it turns out that in most cases CG can be used to solve this equation. If not, we fall back to the CSCG solver to ensure convergence.

In order to find the optimum energy shift μ, we now analyze the convergence behavior of the ARPACK-SI algorithm for a realistic nanostructure. Here, we calculate the lowest electron and hole states for an InAs pyramidal quantum dot using the eight-band $\mathbf{k}\cdot\mathbf{p}$ method. The dot has a width of 15 nm, a height of 7.5 nm, and is embedded in GaAs. The simulation domain is discretized with grid spacings of 0.5 nm and 0.25 nm in the lateral and vertical directions, respectively. This results in a total number of approximately $50,000$ grid nodes within the area where the Schrödinger equation is solved. In Fig. 2.6, we plot the total number of MATMULs (matrix-vector multiplications) that are required for the electron and hole ground states to converge (with an accuracy of 10^{-10} eV), as a function of the relative electron and hole energy shifts $\Delta\mu_e$ and $\Delta\mu_h$, respectively. The absolute energy shifts are given by

$$\mu_e = E_c^{\min} - \Delta\mu_e,\ \mu_h = E_v^{\max} + \Delta\mu_h, \tag{2.41}$$

where $E_c^{\min}$ ($E_v^{\max}$) denotes the global minimum (maximum) of the conduction (valence) band edge in position space. The total number of MATMULs is a good measure for the computational cost, because these are the most time consuming operations. We observe a small decrease of the number of MATMULs when the absolute shift

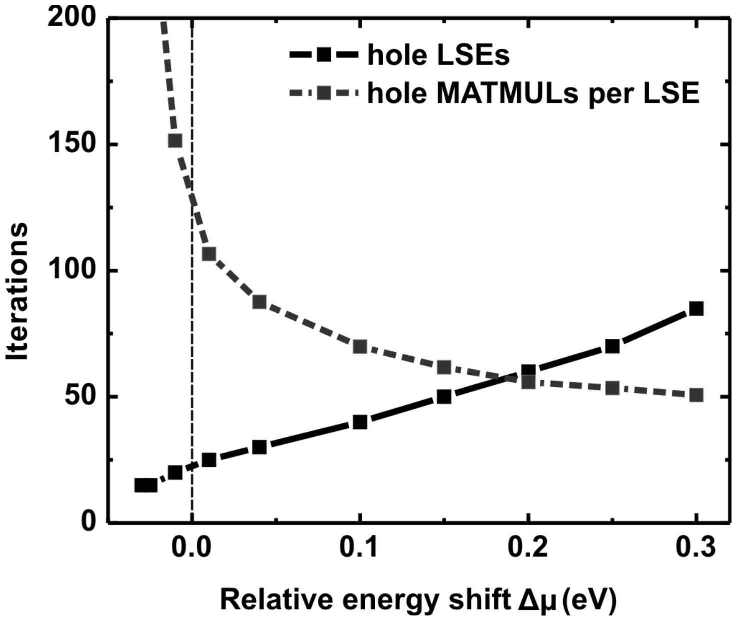


Figure 2.7: Numbers of linear system of equations solutions (solid line) and matrix-vector multiplications per linear equation solution (dashed line) that are required for the calculation of the hole ground state, as a function of the energy shift.

μ is moved from the center of the band gap ($\Delta\mu \approx 0.2$ eV) toward the band edges ($\Delta\mu = 0$). A minimum is reached at approximately $\Delta\mu = 0.05$ eV, followed by a rather sharp increase for values outside the band gap ($\Delta\mu < 0$). This behavior is qualitatively equal for electrons and holes and can be explained by two competing effects. In Fig. 2.7, we show the number of linear systems of equations (LSEs) that have to be solved (i.e. the number of Arnoldi iterations), and the average number of MATMULs per LSE [Eq. (2.40)] for the calculation of the hole ground state, as a function of $\Delta\mu$. When the relative shift $\Delta\mu$ is decreased, the absolute shift μ moves closer to the energies of the lowest hole states and the relevant part of the spectrum is amplified more effectively. Therefore, the number of Arnoldi iterations (LSEs) decreases in order to obtain a converged result for the hole ground state (see solid line in Fig. 2.7). On the other hand, the matrix in Eq. (2.40) becomes singular for $\mu = \lambda$. Thus, the closer one gets to this condition, the more CG iterations (and MATMULs) are required to solve each linear system of equations (see dashed line in Fig. 2.7). These competing effects explain the convergence behavior in Fig. 2.6.

Taking these results into consideration, we have developed the following robust strategy for the determination of the lowest electron and hole states of the eight-band Schrödinger equation with the ARPACK-SI algorithm. We first set the energy shift μ to a value slightly below the upper boundary of the band gap to calculate the electron eigenvalues and eigenstates above the band gap. Afterwards, we set

μ to a value slightly above the lower boundary of the gap to calculate the hole eigenvalues and eigenstates. These energy shifts can be easily estimated in advance from the band edges in position space. The method is reliable even for the calculation of larger numbers of eigenstates. Here, the decreasing effectiveness of the spectral transformation is stabilized by the fact that the CG method (which is used to solve Eq. 2.40) converges faster with increasing energy difference from the singularity. By employing this method in next**nano**++, we are able to calculate the electron and hole ground states of the realistic test structure considered above in approximately 10 minutes on a standard off-the-shelf PC (Intel Core 2, 2.4 GHz). We would like to note that this concept also works for nanostructures that do not have a global energy gap as discussed in chapter 6. In this case, the Schrödinger equation is solved only once with the energy shift set to $(E_c^{\mathrm{min}} + E_v^{\mathrm{max}})/2$.

2.6 Solution of coupled systems of equations

After having provided solution algorithms for all individual equations, the remaining task is to examine how coupled systems can be solved self-consistently. In next**nano**++, these are the coupled Schrödinger- and Poisson equations for equilibrium situations (cf. Sec. 1.3.2) and the coupled Schrödinger-, Poisson-, and current equations for non-equilibrium situations (cf. Sec. 1.5). Since Poisson's equation and the drift-diffusion current equation are boundary value problems and Schrödinger's equation is an eigenvalue problem, there is no obvious way to solve these equations simultaneously, and we have to rely on an iterative approach. We will call such methods "outer iteration" to avoid any confusion with iterations that are needed to solve the individual linear systems of equations and matrix eigenvalue problems. Due to the strong nonlinear couplings between the equations, a straightforward iteration by itself usually does not converge and more sophisticated methods are required to stabilize the convergence. In the following, we will present the particular solution strategies which we use for equilibrium and non-equilibrium situations and that detail the box "Solve coupled equations" in Fig. 2.1.

2.6.1 Coupled Schrödinger-Poisson equations

The basic problem in the system of the Schrödinger equation [Eq. (1.41)]

$$\hat{H}\left[\phi\right]\psi_i = E_i\psi_i \;\; \left(i = 1, ..., N_{ev}\right), \tag{2.42}$$

and the Poison equation [Eq. (1.65)]

$$\boldsymbol{\nabla}\varepsilon\boldsymbol{\nabla}\phi = -\rho\left[\{\psi_i, E_i\}\right], \tag{2.43}$$

is the strong coupling between both equations that is transmitted by the quantum charge density $\rho = \rho_{\mathrm{qm}}$ [Eq. (1.54)] and the electrostatic potential ϕ. The charge

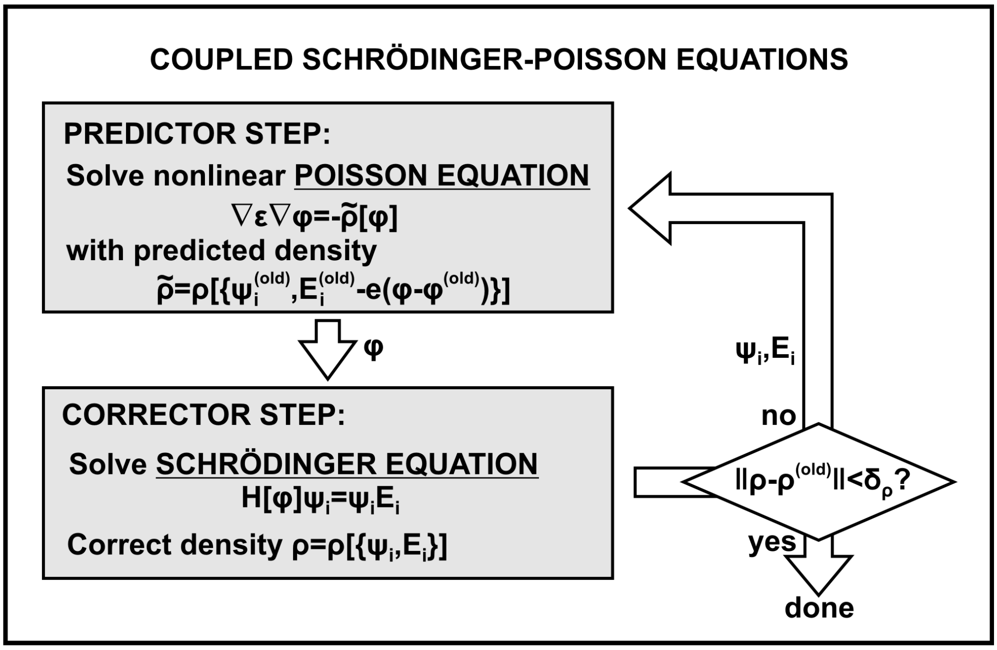


Figure 2.8: Flow scheme of predictor-corrector algorithm for coupled Schrödinger-Poisson equations.

density results from the eigenstates ψ_i and eigenenergies E_i of the Schrödinger equation and enters the Poisson equation, while the electrostatic potential results from the Poisson equation and enters the Hamiltonian of the Schrödinger equation. For this coupled system of equations, the main task is to calculate a self-consistent solution for the charge density (or equivalently for the electrostatic potential). A simple iteration between the Schrödinger and the Poisson equation usually fails to converge due to strong charge oscillations from one iteration step to the other. The most common approach for this problem is to underrelax in the charge density to damp the oscillations [60]. However, the efficiency and stability of this method have remained an issue especially for two- or three-dimensional grids. In next**nano**++, we have improved this situation immensely, by implementing an iteration scheme that is based on a predictor-corrector type approach [61]. Here, the iteration is modified in a way which partially decouples both partial differential equations to stabilize convergence. In Fig. 2.8, we show a flow scheme of this algorithm.

The method is based on the following idea. If we knew the exact dependence of the set of eigenpairs $\{\psi_i, E_i\}$ and therefore the quantum charge density $\rho\left[\{\psi_i, E_i\}\right]$ on the electrostatic potential ϕ, we could solve a nonlinear Poisson equation

$$\boldsymbol{\nabla}\varepsilon\boldsymbol{\nabla}\phi = -\rho\left[\{\psi_i, E_i\}(\phi)\right], \tag{2.44}$$

using Newton's method from Sec. 2.4.2. Then, the Schrödinger equation would be completely decoupled and a single solution would suffice. Since this is not possible, quantum mechanical perturbation theory is used to derive an approximate expression describing the dependence of the quantum charge density on the electrostatic potential. The approximate quantum charge density in the $(k+1)$-th iteration can be obtained by

$$\widetilde{\rho}\left[\left\{\psi_i^{(k)}, E_i^{(k)}\right\}\left(\phi^{(k+1)}\right)\right] = \rho\left[\left\{\psi_i^{(k)}, E_i^{(k)} - e\left(\phi^{(k+1)} - \phi^{(k)}\right)\right\}\right], \tag{2.45}$$

i.e. all eigenenergies of electrons and holes in Eq. (1.54) are shifted according to $E_i^{(k)} \rightarrow E_i^{(k)} - e\left(\phi^{(k+1)} - \phi^{(k)}\right)$ [61]. The density $\widetilde{\rho}$ is used to move most nonlinearities into the Poisson equation

$$\boldsymbol{\nabla}\varepsilon\boldsymbol{\nabla}\phi^{(k+1)} = -\widetilde{\rho}\left[\left\{\psi_i^{(k)}, E_i^{(k)}\right\}\left(\phi^{(k+1)}\right)\right], \tag{2.46}$$

to decrease the coupling in the Schrödinger-Poisson system and improve convergence speed. The predicted result for $\widetilde{\rho}$ is then corrected by an exact solution of the Schrödinger equation

$$\hat{H}\left[\phi^{(k+1)}\right]\psi_i^{(k+1)} = E_i^{(k+1)}\psi_i^{(k+1)}, \tag{2.47}$$

before proceeding with the next outer iteration step. Once the iteration has converged, we have $\phi^{(k+1)} = \phi^{(k)}$ and therefore $\widetilde{\rho} = \rho$, i.e., the predictor yields the correct density. When the Hamiltonian is augmented by exchange correlation potentials (in order to take into account many-particle effects), we use the predicted density $\widetilde{\rho}$ to determine V_{xc} (cf. Appendix B). As exit condition for the iteration, we require that electron as well as hole densities become stationary

$$\begin{aligned} \int d^d\mathbf{x}\left|n^{(k+1)}(\mathbf{x}) - n^{(k)}(\mathbf{x})\right| &< \delta_\rho, \\ \int d^d\mathbf{x}\left|p^{(k+1)}(\mathbf{x}) - p^{(k)}(\mathbf{x})\right| &< \delta_\rho. \end{aligned} \tag{2.48}$$

Reasonable values for the residual δ_ρ are 10^5 cm^{-2}, 10^3 cm^{-1}, and 10^{-3} for one-, two-, and three-dimensional grids ($d = 1, 2, 3$), respectively. In an iterative algorithm, the total number of iterations which are required for convergence always depends on how close the starting value lies to the final result. Primarily, we initialize the potential in a way that guarantees charge neutrality at any spatial position. Here we determine the potential values (that shift the band edges with respect to the Fermi energy to compensate positive and negative charges) using a bisection search [34] for every grid node. Afterwards, we solve the nonlinear Poisson equation with the *classical* charge density ρ_{cl} [Eq. (1.61)], only. The resulting electrostatic potential serves as an initial guess for the coupled Schrödinger-Poisson system. Using this potential, the Schrödinger equation is solved once, before the predictor-corrector procedure in Fig. 2.8 is started. This method has been found to be very reliable and to converge rapidly in a large number of applications.

Recently, we have refined this technique further by projecting the Hamiltonian into the subspace spanned by all already known eigenvectors, and then diagonalizing this small subspace rather than the full Hamiltonian. The basic idea behind this approach is that small changes in the potential ϕ often do not perturb the eigenfunctions of the Hamiltonian $\hat{H}\left[\phi\right]$ significantly, but shift their relative energies, leading to reordering and hybridization of states. Thus, by diagonalizing the subspace Hamiltonian

$$\widetilde{H}_{ij}^{(k+1)} = \left\langle \psi_i^{(k)} \right| \hat{H}\left[\phi^{(k+1)}\right] \left| \psi_j^{(k)} \right\rangle , \tag{2.49}$$

we can obtain a new set of approximate eigenvectors $\widetilde{\psi}_i^{(k+1)}$ that are formed by linear combinations of the eigenvectors $\psi_i^{(k)}$ from the last iteration, and act in place of the exact eigenvectors $\psi_i^{(k+1)}$ of $\hat{H}\left[\phi^{(k+1)}\right]$. This effectively replaces the calculation of N_{ev} eigenvectors of a $N \times N$ matrix by the diagonalization of a $N_{ev} \times N_{ev}$ matrix. Since the number of eigenvectors N_{ev} that contribute to the charge density is rather small (typically $N_{\text{ev}} < 100$), while the number of grid nodes N can be as large as 10^6, this reduces the computational effort drastically. At least in every second outer iteration step, we use an approximate subspace solution of the Schrödinger equation rather than an exact one. Further approximate solutions are inserted if the residuals of the electron and hole charge densities are greater than the convergence criterion δ_ρ times a subspace residual factor f_{subspace} that we usually choose to be 10^6. Basically, this leads to several additional subspace iteration steps at the beginning of the outer iteration process. Since the cpu time spent within the subspace iterations is negligible with respect to the exact iterations, their employment leads to a speed improvement of roughly a factor of two. Note that the predictor-corrector procedure is always terminated by an exact solution of the Schrödinger equation.

2.6.2 Coupled Schrödinger-Poisson-Current equations

Non-equilibrium situations are modeled by locally varying quasi Fermi levels $E_{F,n}\left(\mathbf{x}\right)$ and $E_{F,p}\left(\mathbf{x}\right)$ for electrons and holes, respectively, rather than a single constant energy E_F. In addition to the coupled Schrödinger-Poisson equations from Sec. 2.6.1, we also have to solve the drift-diffusion current equations (1.105)

$$\begin{aligned}
\boldsymbol{\nabla}\mu_n\left(\mathbf{x}\right) n\left(\mathbf{x}\right) \boldsymbol{\nabla} E_{F,n}\left(\mathbf{x}\right) &= R\left(\mathbf{x}\right), \\
\boldsymbol{\nabla}\mu_p\left(\mathbf{x}\right) p\left(\mathbf{x}\right) \boldsymbol{\nabla} E_{F,p}\left(\mathbf{x}\right) &= -R\left(\mathbf{x}\right),
\end{aligned} \tag{2.50}$$

to determine the quasi Fermi energies. These energies become additional variables in the charge densities. So, we have in fact $n\left[\left\{\psi_i, E_i\right\}, E_{F,n}\right]$, $p\left[\left\{\psi_i, E_i\right\}, E_{F,p}\right]$ and $\rho\left[\left\{\psi_i, E_i\right\}, E_{F,n}, E_{F,p}\right] = -n + p$. Therefore, the quasi Fermi levels that result from the current equations now enter the Poisson equation via the charge density and lead to a coupling of both equations. In addition, the current equations are coupled to the Schrödinger equation as the densities depend on the eigenstates and eigenenergies.

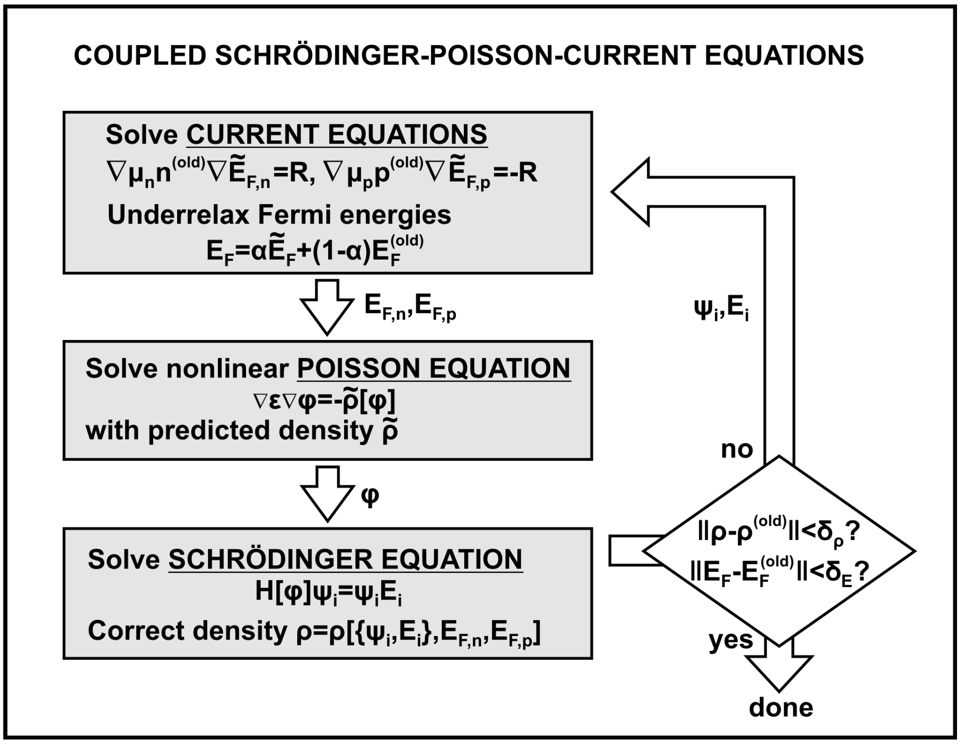


Figure 2.9: Flow scheme of combined predictor-corrector-underrelaxation algorithm for coupled Schrödinger-Poisson-Current equations.

Since there exists no predictor-corrector procedure that includes the current equation, we have to rely on an iterative approach. In Ref. [5] it has been proposed to solve the system of coupled equations in a block-iterative way. This approach consists of alternating solutions of the coupled Schrödinger-Poisson system with fixed quasi Fermi levels $E_{F,n}, E_{F,p}$ and the current equations with fixed eigenpairs $\{\psi_i, E_i\}$. In each of these two blocks, the relevant equations are solved repeatedly until the block result becomes stationary, before proceeding with the next outer iteration step in the other block. This method has been implemented in the previous version next**nano**^{3}. However, we find it to converge only slowly, mainly because it takes up too many Schrödinger solution cycles that are computationally very costly. Calculating a fully converged result for the Schrödinger-Poisson block in each outer iteration step is a waste of cpu time, because in most cases the modifications of the Fermi levels in the current equation strongly perturb the Schrödinger-Poisson solution, anyway. We have improved this method by a novel flow scheme shown in Fig. 2.9 that combines the predictor-corrector procedure for the Schrödinger-Poisson system with an underrelaxation approach for the quasi Fermi levels.

The algorithm can be summarized as follows. We solve the linear current equation for electrons

$$\boldsymbol{\nabla}\mu_n n\left[\left\{\psi_i^{(k)},E_i^{(k)}\right\},E_{F,n}^{(k)}\right]\boldsymbol{\nabla}\widetilde{E}_{F,n}^{(k+1)}=R\left[\left\{\psi_i^{(k)},E_i^{(k)}\right\},E_{F,n}^{(k)}\right],\tag{2.51}$$

and the analogous equation for holes, to obtain the preliminary quasi Fermi energies $\widetilde{E}_{F,n}^{(k+1)}$ and $\widetilde{E}_{F,p}^{(k+1)}$. Afterwards, the new energies $E_{F,n}^{(k+1)}$ and $E_{F,p}^{(k+1)}$ are determined according to

$$E_{F,n}^{(k+1)}=\alpha\widetilde{E}_{F,n}^{(k+1)}+(1-\alpha)\,E_{F,n}^{(k)},\tag{2.52}$$

$$E_{F,p}^{(k+1)}=\alpha\widetilde{E}_{F,p}^{(k+1)}+(1-\alpha)\,E_{F,p}^{(k)},\tag{2.53}$$

with the underrelaxation parameter α. Using these new Fermi energies, only *one* predictor-corrector step is performed before proceeding with the next outer iteration cycle. In this step, we use the predictor density

$$\widetilde{\rho}\left[\phi^{(k+1)}\right]=\rho\left[\left\{\psi_i^{(k)},E_i^{(k)}-e\left(\phi^{(k+1)}-\phi^{(k)}\right)\right\},E_{F,n}^{(k+1)},E_{F,p}^{(k+1)}\right],\tag{2.54}$$

and the corrector density

$$\rho\left[\left\{\psi_i^{(k+1)},E_i^{(k+1)}\right\},E_{F,n}^{(k+1)},E_{F,p}^{(k+1)}\right].\tag{2.55}$$

This procedure turns out to work reliable and to converge reasonably fast in a wide range of applications, when using underrelaxation parameters in the typical range of $\alpha = 0.1 - 0.3$. The major issue that remains, is the fact that charge densities in semiconductor devices can vary by many orders of magnitude on a nanometer scale. Therefore, the matrix that results from discretizing Eq. (2.50) can become very ill-conditioned. This can be even worse in intermediate results of the combined predictor-corrector-underrelaxation algorithm, which need not be physical at all. In order to guarantee convergence of the linear equation [Eq. (2.50)] in every single current calculation step, we limit the spatial variation of charge densities and replace them by a certain minimum value, if necessary. So, in fact, we use the following densities in the current equation,

$$n\left(\mathbf{x}\right)\equiv\max\left[n\left(\mathbf{x}\right),\rho_{\min}\right],\tag{2.56}$$

$$p\left(\mathbf{x}\right)\equiv\max\left[p\left(\mathbf{x}\right),\rho_{\min}\right].\tag{2.57}$$

The minimum density $\rho_{\min}$ should be chosen as large as possible, but smaller than the minimum of the density in the final result (otherwise $\rho_{\min}$ could distort this final result). Typical values for $\rho_{\min}$ lie in the range $10^8 - 10^{12}$ cm^{-3}.

To obtain a good initial guess for the Fermi levels $E_{F,n}$ and $E_{F,p}$, we first solve a linear pseudo current equation

$$\boldsymbol{\nabla}\exp\left[-\frac{E_g\left(\mathbf{x}\right)}{E_0}\right]\boldsymbol{\nabla}E_{F,0}\left(\mathbf{x}\right)=0,\tag{2.58}$$

where $E_g(\mathbf{x})$ is the local band gap and E_0 is some reference energy that can be used to control the Fermi level profile $E_{F,0}(\mathbf{x})$. If E_0 is chosen to be small relative to the range of band gap energies, $E_{F,0}(\mathbf{x})$ will drop in areas with large E_g and remain almost constant in areas of small E_g. This is a good starting value for devices where charges accumulate in small band gap regions that are separated by large band gap barriers. If E_0 is chosen to be large, by contrast, $E_{F,0}(\mathbf{x})$ will drop continuously, almost irrespective on the size of the local energy gap. Typically, we use a value of $E_0 = 1$ eV. Starting with $E_{F,n} = E_{F,p} = E_{F,0}$, we now solve the coupled Poisson-Current system for the *classical* charge density only. Here, the solution of the Schrödinger equation is omitted and the convergence is stabilized by the underrelaxation of the Fermi levels. The resulting quasi Fermi levels and the electrostatic potential serve as initial guess for the coupled Schrödinger-Poisson-Current system. Using the latest potential, the Schrödinger equation is solved once, before the procedure in Fig. 2.9 is started.

As exit condition, we now require separate convergence of electron and hole Fermi levels as well as densities. So in addition to Eq. (2.48), the following relations have to be fulfilled,

$$\begin{aligned} \left\| E_{F,n}^{(k+1)} - E_{F,n}^{(k)} \right\|_\infty &< \delta_E, \\ \left\| E_{F,p}^{(k+1)} - E_{F,p}^{(k)} \right\|_\infty &< \delta_E. \end{aligned} \tag{2.59}$$

Reasonable values for the residual δ_E lie in the range of 10^{-7} to 10^{-5} eV.

Analogous to the pure Schrödinger-Poisson system, the method is combined with the subspace iteration technique to save a considerable amount of exact solution cycles of the Schrödinger equation. In order to assess the convergence behavior of the present method, we investigate an InGaN double quantum well LED that has been specified in Sec. 1.5 (cf. Fig. 1.9). This test structure represents to some extend a worst case scenario for the quantum drift-diffusion (QDD) current model, because the large barrier between the quantum wells strongly limits the current flow and there are pronounced spatial variations of charge densities. In Fig. 2.10, we plot the residuals of the charge densities and the Fermi levels for electrons and holes as a function of the number of exact Schrödinger solution steps. Since these steps are by far the most time consuming ones, they form a good measure for the computational effort. The graph for $\alpha = 1$ shows strong oscillations, which indicate that the convergence fails completely without underrelaxation. The oscillations can be damped with an underrelaxation parameter of $\alpha = 0.2$ leading to reasonably fast convergence. By using the subspace iteration technique, the algorithm exhibits a significant further improvement. Here, we have replaced every second exact solution of the Schrödinger equation by an approximate subspace step. The fast drop of the residual at the beginning comes from several additional subspace iterations that have been inserted between the first and the second exact solution of the Schrödinger equation. These results show that our novel iteration scheme can be applied even for such numerically unfavorable devices. Successful application of the present method to larger quasi three-dimensional systems has been demonstrated in Ref. [62].

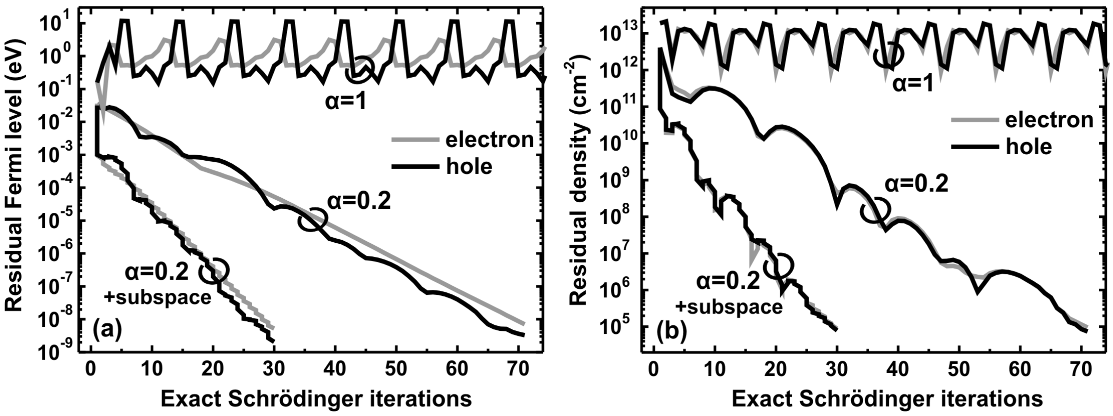


Figure 2.10: Residuals of (a) Fermi levels and (b) charge densities for electrons (gray lines) and holes (black lines) of an $\mathrm{In}_{0.2}\mathrm{Ga}_{0.8}\mathrm{N}$-GaN double quantum well LED as a function of the number of exact Schrödinger solution steps. We compare iteration schemes with ($\alpha = 0.2$) and without ($\alpha = 1$) underralaxation and investigate the effect of subspace iterations.

2.7 Summary

In summary, we have shown how the calculation of semiconductor nanostructures has been realized numerically within next**nano**++. This software has been constructed in a modular way from flow control, application specific, and core numerical libraries. Due to the intensive application of modern C++ object oriented programming techniques, a high degree of code reusage became possible. Furthermore, we have introduced the box discretization technique that has been used to transform all partial differential equations into linear systems of equations or large sparse matrix eigenvalue problems in a flux conserving way. This discretization has been derived in a generalized d-dimensional form so that consistency can always be guaranteed. We have presented iterative algorithms for the solution of the linear algebra problems. Here, we rely on already available software libraries and customize these for our needs. Special emphasis has been put on the non-extremal eigenvalue problem that comes with the eight-band Schrödinger equation. For this particular problem, no satisfactory solution existed so far and we have now employed a robust method based on an ARPACK shift-invert (ARPACK-SI) algorithm. Finally, we have described two different approaches for the solution of coupled systems of equations. Namely, we use a predictor-corrector method for the solution of the coupled Schrödinger-Poisson equations and an underrelaxation approach for the coupled Schrödinger-Poisson-Current system. Both methods have been further improved by a novel subspace iteration technique.

Chapter 3

Gauge-invariant discretization in multiband k·p envelope function theory

3.1 Introduction

Solving the Schrödinger equation in a constant magnetic field within the standard minimal coupling scheme is surprisingly problematic for a confined electronic system. Since the vector potential increases linearly with dimension, it is largest at the boundaries where the wave function is usually assumed to vanish and supposed to be no longer physically relevant. This leads to a pronounced gauge dependence of the eigenfunctions and eigenvalues that is often missed or ignored. On the other hand, magnetic-field related properties such as gyromagnetic factors are essential for controlling the spin degree of freedom of confined carriers in the area of spintronics or quantum computation [1]. It is obviously crucial to develop methods that guarantee manifestly gauge-invariant magnetic-field related solutions of the Schrödinger equation.

Surprisingly few theoretical nonperturbative approaches have been developed so far that focus on mesoscopic semiconductor nanostructures in magnetic fields [63–72]. There are atomistic approaches such as the empirical tight-binding method, where electromagnetic fields can be taken into account via a Peierls-type phase factor [73] in the transfer matrix elements [63–67] and pseudopotential methods, where magnetic fields can be incorporated via a magnetic pseudopotential [68, 69]. Recently, an ab-initio method has been developed to incorporate magnetic fields rigorously [70]. Unfortunately, the latter approaches are practicable only for few-atom systems or crystalline systems with a few dozen atoms per unit cell due to the numerical effort involved.

The envelope function approximation (EFA) [8–19] is the method of choice for handling strained semiconductor structures, including relativistic effects, which extend over hundreds of nanometers. However, most approaches that incorporate magnetic fields within the framework of the EFA treat magnetic fields perturbatively [74, 75]. These approaches typically hold only for subtesla magnetic fields in nanostructures; for a quantum dot of 50 nm diameter, for example, the magnetic length becomes comparable to its size already at 0.3 T. Commonly, the envelope function equation is spatially discretized for the calculation of realistic semiconductor devices. Unfortunately, applying discretization schemes to the minimal coupling Hamiltonian breaks the gauge invariance of the discretized equations [71], and the spectrum depends on the particular gauge choice of the vector potential. Recently, a nonperturbative eight-band EFA method has been developed that is similar to the one presented here [72]. It is also based on the gauge-invariant Wilson loop method [76]. While the details of the method are not specified in the paper, we have been able to reproduce the results in Fig. 2 of Ref. [72] by using an arithmetically averaged symmetrized operator ordering rather than the correct ordering of differential operators that has been established in Refs. [15–19]. The very purpose of the present work is to provide a general, consistent, and rigorous derivation of the method for arbitrary EFA models.

In this chapter, we develop a manifestly gauge-invariant, nonperturbative discretization scheme for the multiband EFA in arbitrary magnetic fields that includes relativistic effects and strain and does not contain any field-dependent fitting parameters. It is based on the concept of gauge covariant derivatives that have been developed in the context of lattice gauge theories originally [76, 77]. We present a complete and concise derivation for the general multiband EFA that includes all derivatives up to second order and thoroughly accounts for their correct placement. For the case of a single band without spin-orbit interaction, the results of the present method are identical to those of the scheme developed in Ref. [71]. For the general multiband case, however, a proper generalization has not been developed so far.

While the present method yields phase factors that multiply the zero-field Hamiltonian matrix elements similar to the Peierls phase factors for the hopping matrix elements in tight-binding theory, there are important differences. First, our method assures that the Hamiltonian reduces to the correct continuum minimal coupling Hamiltonian in the limit of infinitesimal grid spacing. Second, the integration path consists of unambiguous straight-line segments. In tight-binding, on the other hand, the integration paths directly connect two atomic sites [64, 65] of fixed distance.

3.2 Gauge invariance in multiband Schrödinger equations

We start our discussion with the real-space, multiband envelope function Schrödinger equation in a magnetic field [8–10, 12]

$$\widehat{\mathcal{H}}\left(\mathbf{x}\right)\mathbf{F}\left(\mathbf{x}\right)=E\mathbf{F}\left(\mathbf{x}\right), \tag{3.1}$$

where $\mathbf{F}$ is a column vector containing the n components of the (spin-dependent) envelope wave function and

$$\widehat{\mathcal{H}}(\mathbf{x}) = \sum_{ij} K_i \widehat{\mathcal{H}}^{ij}(\mathbf{x}) K_j + \sum_i \left[\widehat{\mathcal{H}}^i(\mathbf{x}) K_i + K_i \widehat{\mathcal{H}}^{i\dagger}(\mathbf{x}) \right] + \widehat{\mathcal{H}}^0(\mathbf{x}), \tag{3.2}$$

is the $n \times n$-matrix of the Hamiltonian operator in the n-band EFA. We assume a d-dimensional structure ($d \geq 2$); the indices $i, j = 1, \ldots, d$ denote the Cartesian components. The momentum operators

$$K_i = -\mathrm{i} D_i = -\mathrm{i} \left[\partial_i + \mathrm{i} \frac{e}{\hbar} A_i(\mathbf{x}) \right], \tag{3.3}$$

can be expressed in terms of the gauge covariant derivatives D_i, and A_i are the Cartesian components of the vector potential. The Hamiltonian matrix $\widehat{\mathcal{H}}(\mathbf{x})$ has been decomposed into $n \times n$-matrices $\widehat{\mathcal{H}}^{ij}$ that contain material-dependent $\mathbf{k}\cdot\mathbf{p}$ parameters and are second-order in the operators K_i, whereas $\widehat{\mathcal{H}}^i$ and $\widehat{\mathcal{H}}^0$ results from the first- and zeroth-order terms, respectively. We will refer to these matrices by $\widehat{\mathcal{H}}^\gamma$, where the index $\gamma \in \{0, i, ij\}$. Concrete implementations of $\widehat{\mathcal{H}}(\mathbf{x})$ are given in Sec. 1.2.1. In the applications that will be presented in chapters 4 and 5, we will always rely on the eight-band Hamiltonian (1.26). Here, the coupling of the carrier's spin to the magnetic field is included via a Zeeman-term [Eq. (1.34)] that contributes to $\widehat{\mathcal{H}}^0$. The matrix $\widehat{\mathcal{H}}^0$ is Hermitian by itself, whereas the individual second-order matrices $\widehat{\mathcal{H}}^{ij}$ only obey the relation $\widehat{\mathcal{H}}^{ij} = (\widehat{\mathcal{H}}^{ji})^\dagger$ which suffices to guarantee $\widehat{\mathcal{H}}$ to be Hermitian. We have used the ordering of the differential operators with respect to the material matrices proposed in Refs. [15–19] that avoid unphysical and spurious solutions of the EFA in heterostructures.

It can easily be shown that Eq. (3.1) is invariant under the gauge transformation

$$\mathbf{F}(\mathbf{x}) \rightarrow \mathbf{F}'(\mathbf{x}) = \mathrm{e}^{-\mathrm{i}\frac{e}{\hbar}\Lambda(\mathbf{x})} \mathbf{F}(\mathbf{x}), \tag{3.4}$$

$$A_i(\mathbf{x}) \rightarrow A_i'(\mathbf{x}) = A_i(\mathbf{x}) + \partial_i \Lambda(\mathbf{x}), \tag{3.5}$$

since $\widehat{\mathcal{H}}$ contains only the covariant derivatives D_i and the magnetic field components B_i. Importantly, it is possible to rewrite the derivatives D_i from Eq. (3.3) into a particular form where the vector potential only enters via a phase factor. To this end, we form the limit [76, 77]

$$D_i \mathbf{F}(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \frac{1}{\varepsilon} \left[U(\mathbf{x} + \boldsymbol{\varepsilon}_i, \mathbf{x}) \mathbf{F}(\mathbf{x} + \boldsymbol{\varepsilon}_i) - U(\mathbf{x}, \mathbf{x}) \mathbf{F}(\mathbf{x}) \right], \tag{3.6}$$

with $\boldsymbol{\varepsilon}_i \equiv \varepsilon \hat{\mathbf{e}}_i$ and $\hat{\mathbf{e}}_i$ denoting the unit vector pointing along the i-direction. We define the so-called connection U,

$$U(\mathbf{x} + \boldsymbol{\varepsilon}_i, \mathbf{x}) = \exp \left[\mathrm{i} \frac{e}{\hbar} \int\limits_{\mathbf{x}}^{\mathbf{x}+\boldsymbol{\varepsilon}_i} \mathbf{A}(\mathbf{x}') \, d\mathbf{x}' \right], \tag{3.7}$$

with the integration path being a straight line from $\mathbf{x}$ to $\mathbf{x} + \boldsymbol{\varepsilon}_i$. Indeed, by inserting Eq. (3.7) into Eq. (3.6), we immediately regain Eq. (3.3). Note that $U(\mathbf{x}, \mathbf{x}) = 1$.

3.3 Gauge-invariant discretization

3.3.1 Prerequisites

Next, we will develop a gauge-invariant spatial discretization for the multiband Schrödinger equation [Eq. (3.1)]. We assume, without loss of generality, the spatial grid to be uniform,[1] to be rectangular, and to be oriented parallel to the Cartesian axes. In addition, we assume the space to be d-dimensional generally and enumerate every grid point by a d-dimensional tuple $\mathbf{m}$ of integers.

First, we discretize Eq. (3.1) for the unproblematic case of a vanishing vector potential $\mathbf{A} = \mathbf{0}$, using a finite difference or box discretization scheme that has been discussed in Sec. 2.3. For example, we may approximate the derivatives ∂_i in Eq. (3.2) by forward finite differences

$$\partial_i \mathbf{F}(\mathbf{x}) \to \delta_i \mathbf{F}(\mathbf{m}) = \frac{1}{\varepsilon}\left[\mathbf{F}(\mathbf{m}+\boldsymbol{\varepsilon}_i) - \mathbf{F}(\mathbf{m})\right], \tag{3.8}$$

where the vector $\boldsymbol{\varepsilon}_i \equiv \varepsilon \hat{\mathbf{e}}_i$ points to the nearest neighbor in the positive Cartesian i-direction and $\varepsilon > 0$. Thus, the discretization of Eq. (3.1) will result in the eigenvalue problem

$$\sum_{\mathbf{n}} \hat{H}_0(\mathbf{m},\mathbf{n})\,\mathbf{F}(\mathbf{n}) = E\mathbf{F}(\mathbf{m}), \tag{3.9}$$

where the sum over $\mathbf{n}$ runs over all grid points. Here, $\hat{H}_0(\mathbf{m},\mathbf{n})$ are the components of the discretized zero-field Hamiltonian operator, and $\mathbf{F}(\mathbf{n})$ denotes the discretized envelope function. The dimension of the Hamiltonian matrix is the product of the total number of grid points times the number of included bands n. For simplicity, we only take into account nearest-neighbor interactions. Since the Hamiltonian includes first and second derivatives, however, the matrix elements $\hat{H}_0(\mathbf{m},\mathbf{n})$ are nonzero both for the $2d$ nearest as well as for the $2d(d-1)$ diagonally adjacent (next-nearest) neighbors. The detailed values of $\hat{H}_0(\mathbf{m},\mathbf{n})$ obviously depend on the chosen discretization method.

We now turn to the nontrivial case of nonzero vector potential $\mathbf{A} \neq \mathbf{0}$. In this case, we must construct suitable discrete approximations Δ_i of the continuous gauge covariant derivatives D_i,

$$D_i \mathbf{F}(\mathbf{x}) \to \Delta_i \mathbf{F}(\mathbf{m}), \tag{3.10}$$

which will lead to the discrete eigenvalue problem

$$\sum_{\mathbf{n}} \hat{H}(\mathbf{m},\mathbf{n})\,\mathbf{F}(\mathbf{n}) = E\mathbf{F}(\mathbf{m}), \tag{3.11}$$

where $\hat{H}(\mathbf{m},\mathbf{n})$ are the Hamiltonian matrix elements in the magnetic field. Importantly, a straightforward discretization of Eq. (3.3) that follows the principle of

---

[1] We note that Eq. (3.11) becomes a generalized eigenvalue problem for nonuniform grids. However, one can apply a scale transformation that leads one back to the form of Eq. (3.11) (cf. Sec. 2.3).

Eq. (3.8),

$$\Delta_i \mathbf{F}(\mathbf{m}) \overset{?}{=} \frac{1}{\varepsilon}\left[\mathbf{F}(\mathbf{m}+\boldsymbol{\varepsilon}_i) - \mathbf{F}(\mathbf{m})\right] + \mathrm{i}\frac{e}{\hbar}A_i(\mathbf{m})\,\mathbf{F}(\mathbf{m}), \tag{3.12}$$

results in a Schrödinger equation that *cannot* be chosen to be gauge-invariant on all grid points, particularly not simultaneously on positions $\mathbf{m}$ and $\mathbf{m}+\boldsymbol{\varepsilon}_i$. Consequently, such a discretization would lead to a spectrum that depends markedly on the chosen gauge for the vector potential $\mathbf{A}$ [71]. We will now show, how to derive a suitable discrete approximation Δ_i that leads to a manifestly covariant eigenvalue problem.

3.3.2 Theorem and corollaries

In the field-free case, we consider the discrete approximations δ_i and δ_{ij} of the first- and second-order derivatives ∂_i and $\partial_i\partial_j$, respectively, which can be written in the following general form,

$$\begin{aligned}
\delta_i \mathbf{F}(\mathbf{m}) &= \frac{1}{\varepsilon}\sum_{s_i} C_i(s_i)\,\mathbf{F}(\mathbf{m}+s_i\boldsymbol{\varepsilon}_i), \\
\delta_{ii} \mathbf{F}(\mathbf{m}) &= \frac{1}{\varepsilon^2}\sum_{s_i} C_{ii}(s_i)\,\mathbf{F}(\mathbf{m}+s_i\boldsymbol{\varepsilon}_i), \\
\delta_{ij} \mathbf{F}(\mathbf{m}) &= \frac{1}{\varepsilon^2}\sum_{s_i,s_j} C_{ij}(s_i,s_j)\,\mathbf{F}(\mathbf{m}+s_i\boldsymbol{\varepsilon}_i+s_j\boldsymbol{\varepsilon}_j),
\end{aligned} \tag{3.13}$$

where $i \neq j$ and $s_i, s_j \in \{0, \pm 1\}$, so that the sums run over the grid point $\mathbf{m}$ and its neighbors. The coefficients C have to be chosen in such a way that they guarantee the limits $(i, j = 1, ..., d)$.

$$\lim_{\varepsilon\to 0} \delta_i \mathbf{F}(\mathbf{m}) = \partial_i \mathbf{F}(\mathbf{x}), \tag{3.14}$$

$$\lim_{\varepsilon\to 0} \delta_{ij} \mathbf{F}(\mathbf{m}) = \partial_i\partial_j \mathbf{F}(\mathbf{x}). \tag{3.15}$$

Concretely, we have used the following values for the discretization coefficients in the applications that we will present in chapter 4,

$$\begin{aligned}
&C_i(1) = 1, C_i(0) = -1, C_i(-1) = 0, \\
&C_{ii}(0) = -2, C_{ii}(\pm 1) = 1, C_{ij}(s_i, s_j) = \frac{1}{4}s_i s_j.
\end{aligned} \tag{3.16}$$

They can be obtained from Eqs. (2.11), (2.13), and (2.16) assuming a homogenous material composition as well as a uniform grid. Additionally, we may split the matrix elements $\hat{H}_0(\mathbf{m},\mathbf{n})$ of the discretized field-free Hamiltonian in Eq. (3.9) into their contributions corresponding to the zeroth, first and second derivatives $\widehat{\mathcal{H}}^{\gamma}$ in Eq. (3.2),

$$\hat{H}_0(\mathbf{m},\mathbf{n}) = \sum_{\gamma} \hat{H}_0(\mathbf{m},\mathbf{n},\gamma), \tag{3.17}$$

with $\gamma \in \{0, i, ij\}$.

Definition. We define the discrete first-order approximation Δ_i of the gauge covariant derivative D_i and the second-order approximations Δ_{ii} and Δ_{ij} of D_i^2 and D_iD_j, respectively, by multiplying each term in Eq. (3.13) by the discrete connection $U\left(\mathbf{m},\mathbf{n},\gamma\right)$,

$$\begin{aligned}
\Delta_i\mathbf{F}\left(\mathbf{m}\right) &= \frac{1}{\varepsilon}\sum_{s_i} C_i\left(s_i\right)U\left(\mathbf{m},\mathbf{m}+s_i\boldsymbol{\varepsilon}_i,i\right)\mathbf{F}\left(\mathbf{m}+s_i\boldsymbol{\varepsilon}_i\right),\\
\Delta_{ii}\mathbf{F}\left(\mathbf{m}\right) &= \frac{1}{\varepsilon^2}\sum_{s_i} C_{ii}\left(s_i\right)U\left(\mathbf{m},\mathbf{m}+s_i\boldsymbol{\varepsilon}_i,i\right)\mathbf{F}\left(\mathbf{m}+s_i\boldsymbol{\varepsilon}_i\right),\\
\Delta_{ij}\mathbf{F}\left(\mathbf{m}\right) &= \frac{1}{\varepsilon^2}\sum_{s_i,s_j} C_{ij}\left(s_i,s_j\right)U\left(\mathbf{m},\mathbf{m}+s_i\boldsymbol{\varepsilon}_i+s_j\boldsymbol{\varepsilon}_j,ij\right)\mathbf{F}\left(\mathbf{m}+s_i\boldsymbol{\varepsilon}_i+s_j\boldsymbol{\varepsilon}_j\right),
\end{aligned}\tag{3.18}$$

where the last line holds for $i \neq j$. The discrete connection U is defined by

$$U\left(\mathbf{m},\mathbf{n},\gamma\right) = \exp\left[\mathrm{i}\frac{e}{\hbar}\int\limits_{S(\mathbf{m},\mathbf{n},\gamma)}\mathbf{A}\left(\mathbf{x}'\right)d\mathbf{x}'\right],\tag{3.19}$$

where the trajectory $S\left(\mathbf{m},\mathbf{n},\gamma\right)$ starts at position $\mathbf{m}$ and ends at one of the nearest- or next-nearest-neighbor points $\mathbf{n}$. The specifier γ labels the concrete straight-line segments along the Cartesian axes i,j of a particular trajectory as follows,

$$\begin{aligned}
\gamma = i,&\ \text{Path: } \mathbf{m}\rightarrow\mathbf{m}\pm\boldsymbol{\varepsilon}_i=\mathbf{n},\\
\gamma = ij,&\ \text{Path: } \mathbf{m}\rightarrow\mathbf{m}\pm\boldsymbol{\varepsilon}_i=\mathbf{l}\rightarrow\mathbf{l}\pm\boldsymbol{\varepsilon}_j=\mathbf{n},\quad (i\neq j)\\
\gamma^\dagger = ji,&\ \text{Path: } \mathbf{m}\rightarrow\mathbf{m}\pm\boldsymbol{\varepsilon}_j=\mathbf{l}'\rightarrow\mathbf{l}'\pm\boldsymbol{\varepsilon}_i=\mathbf{n}.\quad (i\neq j)
\end{aligned}\tag{3.20}$$

Here, each segment $\boldsymbol{\varepsilon}_i$ connects only nearest neighbors as indicated in Fig. 3.1. The choice of this integration path guarantees that the canonical momenta obey the well-known commutator relation for finite magnetic fields. As we will show below, this requires that the trajectories associated with $\gamma = ij$ and $\gamma^\dagger = ji$, enclose an area ϵ^2 that is compatible with the discretization stencil.

Theorem. The expressions of Eq. (3.18) guarantee the correct continuum limits

$$\lim_{\varepsilon\rightarrow 0}\Delta_i\mathbf{F}\left(\mathbf{m}\right) = D_i\mathbf{F}\left(\mathbf{x}\right),\tag{3.21}$$

$$\lim_{\varepsilon\rightarrow 0}\Delta_{ij}\mathbf{F}\left(\mathbf{m}\right) = D_iD_j\mathbf{F}\left(\mathbf{x}\right).\tag{3.22}$$

Note that one has $D_iD_j \neq D_jD_i$ for $i \neq j$.

Corollary 1. The matrix elements of the discretized Hamiltonian $\hat{H}$ for nonzero vector potential $\mathbf{A} \neq \mathbf{0}$ can be written in the form

$$\hat{H}\left(\mathbf{m},\mathbf{n},\gamma\right) = \hat{H}_0\left(\mathbf{m},\mathbf{n},\gamma\right)U\left(\mathbf{m},\mathbf{n},\gamma\right),\tag{3.23}$$

$$U\left(\mathbf{m},\mathbf{m},\gamma\right) = U(\mathbf{m},\mathbf{m}) = 1,\tag{3.24}$$

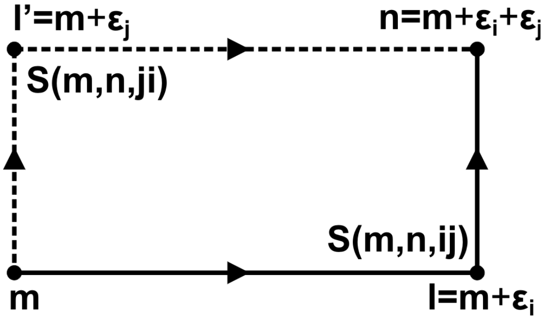


Figure 3.1: Alternative integration paths in the discretization of mixed second order derivatives.

so that the discretized Schrödinger equation [Eq. (3.11)] reads

$$\sum_{\mathbf{n},\gamma} \hat{H}_0\left(\mathbf{m},\mathbf{n},\gamma\right) U\left(\mathbf{m},\mathbf{n},\gamma\right) \mathbf{F}\left(\mathbf{n}\right) = E\mathbf{F}\left(\mathbf{m}\right). \tag{3.25}$$

Corollary 2. The discrete connection [Eq. (3.19)] obeys the following condition

$$U^*\left(\mathbf{n},\mathbf{m},\gamma^\dagger\right) = U\left(\mathbf{m},\mathbf{n},\gamma\right), \tag{3.26}$$

which guarantees the Hermiticity of the Hamiltonian $\hat{H}$.

Corollary 3. The discrete connection [Eq. (3.19)] obeys the following discrete gauge transformation,

$$U\left(\mathbf{m},\mathbf{n},\gamma\right) \to U'\left(\mathbf{m},\mathbf{n},\gamma\right) = \exp\left[-\mathrm{i}\frac{e}{\hbar}\Lambda\left(\mathbf{m}\right)\right] U\left(\mathbf{m},\mathbf{n},\gamma\right) \exp\left[\mathrm{i}\frac{e}{\hbar}\Lambda\left(\mathbf{n}\right)\right], \tag{3.27}$$

which guarantees the gauge invariance of the Hamiltonian $\hat{H}$.

Corollary 4. The covariant derivatives fulfill the well known commutator relations in the continuum limit

$$\lim_{\epsilon\to 0}\left[\Delta_x, \Delta_y\right] = \left[D_x, D_y\right] = \mathrm{i}\frac{e}{\hbar}B_z, \tag{3.28}$$

and cyclic permutations.

3.3.3 Proofs

To prove the theorem, we first evaluate the connections U explicitly. For each path segment $\boldsymbol{\varepsilon}_i$ connecting the neighbors $\mathbf{m}$, $\mathbf{n}$ via a path specified by γ,we determine the curve integral in Eq. (3.19) by the trapezoidal rule,

$$\int\limits_{S(\mathbf{m},\mathbf{m}+\boldsymbol{\varepsilon}_i,i)} \mathbf{A}\left(\mathbf{x}'\right) d\mathbf{x}' = \frac{1}{2}\left[A_i\left(\mathbf{m}\right) + A_i\left(\mathbf{m}+\boldsymbol{\varepsilon}_i\right)\right]\varepsilon \equiv \bar{A}_i\left(\mathbf{m}+\boldsymbol{\varepsilon}_i/2\right)\varepsilon. \tag{3.29}$$

In the case of nearest-neighbors $\mathbf{m}$ and $\mathbf{m} + s_i \boldsymbol{\varepsilon}_i$ ($s_i = \pm 1$), this results in the connection

$$U(\mathbf{m}, \mathbf{m} + s_i \boldsymbol{\varepsilon}_i, i) = \exp\left[s_i \mathrm{i} \frac{e}{\hbar} \bar{A}_i(\mathbf{m} + s_i \boldsymbol{\varepsilon}_i/2)\, \varepsilon\right]. \tag{3.30}$$

For the paths connecting next-nearest-neighbors, specified by $\gamma = ij$, we add up the individual straight-line components of the curve integral. With the definition [Eq. (3.19)], this leads to

$$U(\mathbf{m}, \mathbf{m} + s_i \boldsymbol{\varepsilon}_i + s_j \boldsymbol{\varepsilon}_j, ij) = \exp\left[s_i \mathrm{i} \frac{e}{\hbar} \bar{A}_i(\mathbf{m} + s_i \boldsymbol{\varepsilon}_i/2)\, \varepsilon + s_j \mathrm{i} \frac{e}{\hbar} \bar{A}_j(\mathbf{m} + s_i \boldsymbol{\varepsilon}_i + s_j \boldsymbol{\varepsilon}_j/2)\, \varepsilon\right], \tag{3.31}$$

with $s_i, s_j \in \{\pm 1\}$. To show that the discretized Schrödinger equation tends to the continuum case in the limit of $\varepsilon \longrightarrow 0$, we expand the connections U into powers of ε,

$$U(\mathbf{m}, \mathbf{m} + s_i \boldsymbol{\varepsilon}_i, i) = 1 + s_i \mathrm{i} \frac{e}{\hbar} A_i(\mathbf{m})\, \varepsilon + \mathcal{O}\left(\varepsilon^2\right). \tag{3.32}$$

For the wave functions $\mathbf{F}$, we have

$$\mathbf{F}(\mathbf{m} + s_i \boldsymbol{\varepsilon}_i) = \mathbf{F}(\mathbf{m}) + s_i \partial_i \mathbf{F}(\mathbf{m}) \varepsilon + \mathcal{O}\left(\varepsilon^2\right). \tag{3.33}$$

By inserting these relations into the first derivatives [Eq. (3.18)], we obtain

$$\Delta_i \mathbf{F}(\mathbf{m}) = \frac{1}{\varepsilon} \sum_{s_i} C_i(s_i) \left[\mathbf{F}(\mathbf{m}) + s_i \partial_i \mathbf{F}(\mathbf{m}) \varepsilon + s_i \mathrm{i} \frac{e}{\hbar} A_i(\mathbf{m})\, \mathbf{F}(\mathbf{m})\, \varepsilon\right] + \mathcal{O}(\varepsilon). \tag{3.34}$$

Since for $\mathbf{A} = 0$, Eq. (3.34) must reduce to the partial derivative $\partial_i \mathbf{F}(\mathbf{m})$, the coefficients C_i have to fulfill the relations

$$\sum_{s_i} C_i(s_i) = 0, \ \sum_{s_i} C_i(s_i)\, s_i = 1. \tag{3.35}$$

This can be used to simplify the remaining term in Eq. (3.34) for $\mathbf{A} \neq 0$, and leads to the required limit

$$\Delta_i \mathbf{F}(\mathbf{m}) = \partial_i \mathbf{F}(\mathbf{m}) + \mathrm{i} \frac{e}{\hbar} A_i(\mathbf{m})\, \mathbf{F}(\mathbf{m}) + \mathcal{O}(\varepsilon). \tag{3.36}$$

By following arguments along these lines for the second derivatives as well, we finally obtain (see Appendix E for a detailed derivation)

$$\Delta_{ij} \mathbf{F}(\mathbf{m}) = \left[\partial_i + \mathrm{i} \frac{e}{\hbar} A_i(\mathbf{m})\right] \left[\partial_j + \mathrm{i} \frac{e}{\hbar} A_j(\mathbf{m})\right] \mathbf{F}(\mathbf{m}) + \mathcal{O}(\varepsilon), \tag{3.37}$$

which has the correct continuum limit.

Next, we will prove *corollary 1.* This result follows directly from the fact that each zero-field Hamiltonian matrix element $\hat{H}_0(\mathbf{m}, \mathbf{n}, \gamma)$ gets augmented by a corresponding phase factor $U(\mathbf{m}, \mathbf{n}, \gamma)$, $\gamma \in \{0, i, ij\}$, according to Eqs. (3.18). In order

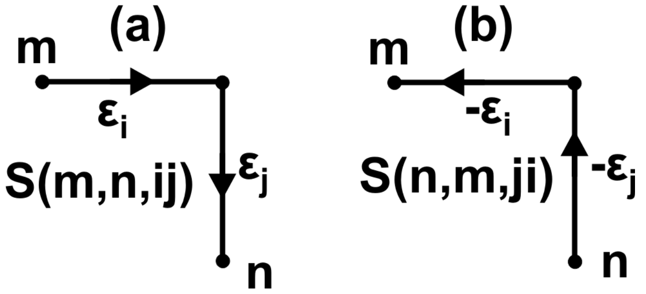


Figure 3.2: Illustration of the integration paths associated with (a) the connection $U(\mathbf{m},\mathbf{n},ij)$ and (b) its Hermitian conjugate $U(\mathbf{n},\mathbf{m},ji)$.

to prove *corollary 2*, we need to show that the Hamiltonian $\hat{H}$ is Hermitian. This implies the relation

$$U^*\left(\mathbf{n},\mathbf{m},\gamma^\dagger\right)\hat{H}_0^\dagger\left(\mathbf{n},\mathbf{m},\gamma^\dagger\right)=\hat{H}^\dagger\left(\mathbf{n},\mathbf{m},\gamma^\dagger\right)\equiv\hat{H}\left(\mathbf{m},\mathbf{n},\gamma\right)=\hat{H}_0\left(\mathbf{m},\mathbf{n},\gamma\right)U\left(\mathbf{m},\mathbf{n},\gamma\right). \tag{3.38}$$

The field-free Hamiltonian is definitely Hermitian and obeys $\hat{H}_0^\dagger(\mathbf{m},\mathbf{n},\gamma^\dagger)=\hat{H}_0(\mathbf{m},\mathbf{n},\gamma)$, since $\widehat{\mathcal{H}}^{ij}=(\widehat{\mathcal{H}}^{ji})^\dagger$. Thus, we are led to the condition

$$U^*\left(\mathbf{n},\mathbf{m},\gamma^\dagger\right)=U\left(\mathbf{m},\mathbf{n},\gamma\right). \tag{3.39}$$

We now show that this equation is fulfilled indeed. For a nearest-neighbor connection, we immediately find this relation to hold,

$$U\left(\mathbf{m},\mathbf{m}+\boldsymbol{\varepsilon}_i,i\right)=\exp\left[-\mathrm{i}\frac{e}{\hbar}\int\limits_{S(\mathbf{m}+\boldsymbol{\varepsilon}_i,\mathbf{m},i)}\mathbf{A}\left(\mathbf{x}'\right)d\mathbf{x}'\right]=U^*\left(\mathbf{m}+\boldsymbol{\varepsilon}_i,\mathbf{m},i\right). \tag{3.40}$$

For the next-nearest-neighbor connection between $\mathbf{m}$ and $\mathbf{n}=\mathbf{m}+\boldsymbol{\varepsilon}_i+\boldsymbol{\varepsilon}_j$ specified by $\gamma=ij$, we obtain

$$\begin{aligned}U\left(\mathbf{m},\mathbf{n},ij\right)&=\exp\left[\mathrm{i}\frac{e}{\hbar}\bar{A}_i\left(\mathbf{m}+\boldsymbol{\varepsilon}_i/2\right)\varepsilon+\mathrm{i}\frac{e}{\hbar}\bar{A}_j\left(\mathbf{m}+\boldsymbol{\varepsilon}_i+\boldsymbol{\varepsilon}_j/2\right)\varepsilon\right]\\&=\exp\left[-\mathrm{i}\frac{e}{\hbar}\bar{A}_j\left(\mathbf{n}-\boldsymbol{\varepsilon}_j/2\right)\varepsilon-\mathrm{i}\frac{e}{\hbar}\bar{A}_i\left(\mathbf{n}-\boldsymbol{\varepsilon}_j-\boldsymbol{\varepsilon}_i/2\right)\varepsilon\right]^*\\&=U^*\left(\mathbf{n},\mathbf{m},ji\right),\end{aligned} \tag{3.41}$$

which is precisely the condition [Eq. (3.39)]. Note that the appropriate paths $S(\mathbf{m},\mathbf{n},ij)$ and $S(\mathbf{n},\mathbf{m},ji)$ contain the same segments but are traversed in the opposite direction (see Fig. 3.2). We now turn to *corollary 3*. Let us write the discretized Schrödinger equation [Eq. (3.25)] in a different gauge,

$$\sum_{\mathbf{n},\gamma}\hat{H}_0\left(\mathbf{m},\mathbf{n},\gamma\right)U'\left(\mathbf{m},\mathbf{n},\gamma\right)\mathbf{F}'\left(\mathbf{n}\right)=E\mathbf{F}'\left(\mathbf{m}\right). \tag{3.42}$$

The relation between the envelope function $F(\mathbf{n})$ and its corresponding form in the primed gauge is the same as in the continuous case and reads

$$\mathbf{F}(\mathbf{n}) \rightarrow \mathbf{F}'(\mathbf{n}) = \exp\left[-\mathrm{i}\frac{e}{\hbar}\Lambda(\mathbf{n})\right]\mathbf{F}(\mathbf{n}). \tag{3.43}$$

Inserting Eq. (3.43) into Eq. (3.42) leads to the requirement that $U(\mathbf{m},\mathbf{n},\gamma)$ has to fulfill a discretized version of the continuum gauge transformation,

$$U(\mathbf{m},\mathbf{n},\gamma) \rightarrow U'(\mathbf{m},\mathbf{n},\gamma) = \exp\left[-\mathrm{i}\frac{e}{\hbar}\Lambda(\mathbf{m})\right]U(\mathbf{m},\mathbf{n},\gamma)\exp\left[\mathrm{i}\frac{e}{\hbar}\Lambda(\mathbf{n})\right]. \tag{3.44}$$

We now show that this condition is fulfilled indeed. The discrete gauge transformation for Eq. (3.29) reads

$$\bar{A}_i'(\mathbf{m}+\boldsymbol{\varepsilon}_i/2) = \bar{A}_i(\mathbf{m}+\boldsymbol{\varepsilon}_i/2) + \frac{1}{\varepsilon}\left[\Lambda(\mathbf{m}+\boldsymbol{\varepsilon}_i) - \Lambda(\mathbf{m})\right]. \tag{3.45}$$

By inserting this expression into the nearest-neighbor connection [Eq. (3.30)], we immediately obtain Eq. (3.44). For the connections to next-nearest-neighbors, we get

$$\begin{aligned} & U'(\mathbf{m},\mathbf{m}+\boldsymbol{\varepsilon}_i+\boldsymbol{\varepsilon}_j,ij) \\ & = \exp\left\{\mathrm{i}\frac{e}{\hbar}\left[\bar{A}_i(\mathbf{m}+\boldsymbol{\varepsilon}_i/2)\,\varepsilon + \Lambda(\mathbf{m}+\boldsymbol{\varepsilon}_i) - \Lambda(\mathbf{m})\right]\right\} \\ & \times \exp\left\{\mathrm{i}\frac{e}{\hbar}\left[\bar{A}_j(\mathbf{m}+\boldsymbol{\varepsilon}_i+\boldsymbol{\varepsilon}_j/2)\,\varepsilon + \Lambda(\mathbf{m}+\boldsymbol{\varepsilon}_i+\boldsymbol{\varepsilon}_j) - \Lambda(\mathbf{m}+\boldsymbol{\varepsilon}_i)\right]\right\} \qquad (3.46) \\ & = \exp\left[-i\frac{e}{\hbar}\Lambda(\mathbf{m})\right]U(\mathbf{m},\mathbf{m}+\boldsymbol{\varepsilon}_i+\boldsymbol{\varepsilon}_j,ij)\exp\left[i\frac{e}{\hbar}\Lambda(\mathbf{m}+\boldsymbol{\varepsilon}_i+\boldsymbol{\varepsilon}_j)\right], \qquad (3.47) \end{aligned}$$

which confirms Eq. (3.44) as well. In addition, the discrete gauge transformation [Eq. (3.45)] obviously has the correct continuum limit [Eq. (3.5)]. Together, these results confirm that the discretized Schrödinger equation [Eq. (3.25)] is invariant under local phase transformations. Finally, we prove *corollary 4*. Using Eq. (3.37), we can see that the discretized covariant derivatives obey the commutator relation [Eq. (3.28)] in the continuum limit,

$$\lim_{\varepsilon\to 0}\left[\Delta_i,\Delta_j\right] = \lim_{\varepsilon\to 0}\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_i A_j - \partial_j A_i\right) + \mathcal{O}(\varepsilon)\right] = \mathrm{i}\frac{e}{\hbar}\varepsilon_{ijk}B_k = \left[D_i, D_j\right]. \tag{3.48}$$

An issue we have not discussed so far concerns the uniqueness of the integration trajectory $S(\mathbf{m},\mathbf{n},\gamma)$. For connections between nearest-neighbors, the integration path is unique. For diagonally adjacent neighbors, however, there are two possible paths to choose from (cf. Fig. 3.1). Gauge invariance only demands that the integration path continuously connects position $\mathbf{m}$ with position $\mathbf{n}$. The Hermiticity of H is guaranteed once the same path is traversed in opposite direction for $H^\dagger$ (cf. Fig. 3.2). In addition, however, the covariant derivatives have to fulfill the commutator relations [Eq. (3.28)]. For the second-order approximations Δ_{ij} in Eq. (3.18), the only element that depends on the order of i and j is the path characterized by $\gamma = ij$, since $C_{ij} = C_{ji}$. Thus, we can see that it is the precise choice of γ in Eq. (3.20) that guarantees the correct order in Eq. (3.37) which in turn guarantees the validity of *corollary 4*. This additional constraint causes the integration trajectory to be unique.

3.4 Summary of method

In summary, we have developed a general method to solve the multi-band Schrödinger equation in an external potential plus a magnetic field in an efficient and manifestly gauge-invariant manner. It is based on the concept of gauge covariant derivatives [76, 77]. The phase factors comprise a discretized form of the curve integral over the vector potential. The integration paths are compatible with the discretization grid and are given by a chain of straight-line path segments connecting nearest neighbors. Their choice is uniquely defined by the constraint that the momentum operators must obey well-known commutator relations. The Hamiltonian matrix elements are shown to be gauge-invariant for any finite grid spacing and converge to the same correct continuum result in the limit of infinitesimal grid spacing. The coupling of the carrier's spin to the magnetic field is properly taken into account by a Zeeman term within the EFA scheme.

To apply this method, one first discretizes the Hamiltonian (3.2) in the field-free Schrödinger equation [Eq. (3.1)] using the discretization scheme of Eq. (3.13) with the coefficients [Eq. (3.16)]. This yields the discretized Schrödinger equation (3.9) with the Hamiltonian matrix consisting of the zeroth-, first- and second-order derivatives [Eq. (3.17)]. Next, each matrix element of the Hamiltonian $\hat{H}_0(\mathbf{m}, \mathbf{n}, \gamma)$ is to be multiplied by the phase factor $U(\mathbf{m}, \mathbf{n}, \gamma)$, in order to take into account the vector potential $\mathbf{A}$. For nearest neighbors, this phase factor is given by Eq. (3.30). For diagonally adjacent neighbors, Eq. (3.31) is needed. In the latter case, one must be careful to discriminate terms of the form $\gamma = ij$ from those that contain $\gamma = ji$. Finally, one needs to add a zeroth-order derivative term, the Zeeman term namely [Eq. (1.34)], to the Hamiltonian in order to include the coupling of the carrier's spin to the magnetic field.

Chapter 4

Nanostructures in high magnetic fields and g factor engineering

4.1 Introduction

During the last few years, there has been an increasing interest in studying the properties of low-dimensional semiconductor nanostructures such as quantum dots and quantum wires in magnetic fields. While it is well known that in bulk semiconductors that are subject to a magnetic field, the energy bands get split into Landau levels, this situation alters significantly in nanostructures. Here, confinement modifies the energy spectrum into quantized levels already at zero magnetic field. Understanding the interplay between the effects of spatial confinement and magnetic fields is far from trivial. In addition, the use of magnetic fields for controlling the spin degree of freedom of confined carriers plays a prominent role in quantum computation and spintronics [1]. Here, it is important to understand the mechanisms that determine magnetic-field related properties such as gyromagnetic factors. These g factors characterize the splitting of spin states in magnetic fields and their spin precession frequency. The latter aspect can be exploited for coherent spin manipulations. Effective g factors have been investigated experimentally in quantum wells [78–81], quantum wires [82], and quantum dots [83–87], and show a strong dependence on the spatial extension, strain and material composition of the system. Hence, the control of this quantity, e.g., by tayloring the size of the nanostructure, enables the fabrication of devices with individually addressable single carrier spins such as spin-qubits.

In this chapter, we investigate the electronic structure of realistic nanostructures in magnetic fields. Here, we rely on our novel nonperturbative method that has been introduced in chapter 3. We start our discussion by giving an overview of the energy spectrum of nanostructures in high magnetic fields, and the mechanisms that determine g tensors in bulk semiconductors and low dimensional nanostructures. In order to critically assess the accuracy of our method, we calculate effective electron and hole g tensors of InP/InAs nanowire-based quantum dots and compare the results

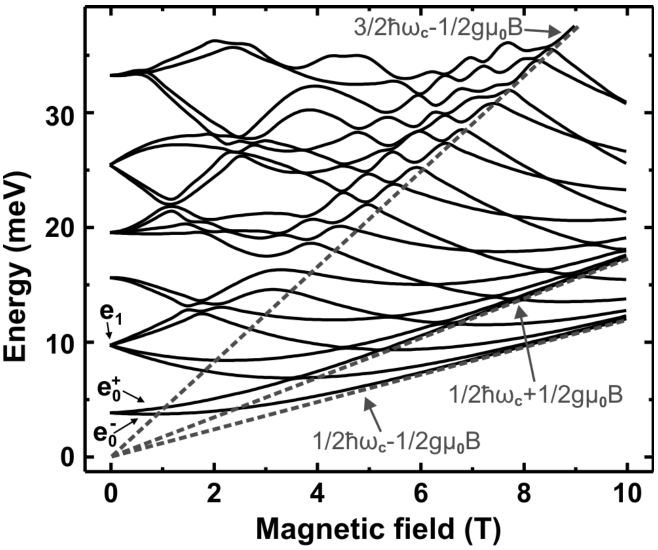


Figure 4.1: Calculated zone center conduction subband energies (full lines) of a 70 nm wide rectangular GaSb quantum wire, as a function of the magnetic field B along the [001] wire axis. The graph shows a continuous transition between the confinement dominated (small B) regime with discrete energy levels and the magnetic field dominated (large B) regime indicated by the Landau level energy fan (dashed lines).

with experimental data [86]. In addition, we discuss the influence of strain and high magnetic fields on the g factors. Finally, we predict electron and hole g factors of self-assembled InAs/GaAs quantum dots as a function of the dot dimensions and external electric fields.

4.2 Energy spectrum and g tensors

4.2.1 Modification of energy spectrum by magnetic fields

We start with a brief discussion of magnetic-field induced changes of the energy spectrum of confined states in semiconductor nanostructures. As an illustration, in Fig. 4.1, we have calculated the lowest conduction subband states at $k_\parallel = 0$ of a 70 nm wide rectangular GaSb quantum wire, as a function of the magnetic field B along the [001] wire axis. The graph shows a spin resolved energy spectrum similar to the Fock-Darwin spectrum, which can be obtained analytically for a two-dimensional parabolic confinement potential with a perpendicular magnetic field [88]. At zero magnetic field, the s-type ground state (e_0) is twofold spin degenerate, while the first excited state (e_1) with a p-type envelope function is fourfold degenerate due to spin and angular momentum. For nonzero magnetic field, these degeneracies are lifted by the Zeeman splitting. In the limit of high magnetic field values, the confinement plays a negligible role compared to the magnetic field effect. Therefore, the eigenenergies gather into a Landau level like energy fan $E_n = (n + 1/2)\hbar\omega_c \pm 1/2\mathrm{g}\mu_0 B$, where ω_c denotes the cyclotron frequency, n is the Landau index, and g is the conduction band g factor (see dashed lines in Fig. 4.1). This demonstrates that low dimensional nanostructures can exhibit a transition between confinement dominated and magnetic-field dominated regimes. We now turn to effects that can be obtained already at small values of B. Further high magnetic field effects will be discussed in Sec. 4.3.4.

4.2.2 Mechanisms of g tensor modulation

Crucial magnetic-field related properties of nanostructures are g factors which characterize the Zeeman splittings of confined eigenstates. In a semiconductor, the g factor differs from the pure spin value of $\mathrm{g}_0 = 2$ due to spin-orbit coupling. In a homogenous bulk material, this can be understood as follows. When a magnetic field is applied, the wave functions are modified into Landau levels, corresponding to quantized orbital angular momentum along the axis of the magnetic field. The Zeeman energy now splits each Landau level into two spin-polarized levels, one with spin parallel to the quantization axis and one antiparallel. Although the pure $\mathrm{g}_0 = 2$ raises the energy of the parallel spin state and lowers that of the antiparallel state, the spin-orbit interaction may favour the opposite spin alignment relative to the orbital angular momentum. When this effect is absorbed into the g factor, it makes $\mathrm{g} < 2$.

Since the lowest conduction band has an s-type Bloch basis with zero orbital momentum, it does not obtain any direct spin-orbit interaction. Here, any deviation from 2 is induced only via the coupling to other bands. In $\mathbf{k}\cdot\mathbf{p}$ theory, these contributions can be derived by a careful treatment of the ordering of momentum matrix elements that do not commute in a magnetic field. Accordingly, in a bulk semiconductor, the conduction band edge g factor is given by

$$\mathrm{g}_c = \mathrm{g}_0 \left(1 - \frac{\mathrm{i}}{m_0} \sum_{\beta \neq c} \frac{\langle u_c | p_x | u_\beta \rangle \langle u_\beta | p_y | u_c \rangle - \langle u_c | p_y | u_\beta \rangle \langle u_\beta | p_x | u_c \rangle}{E_c - E_\beta} \right), \tag{4.1}$$

where the summation runs over all bands, excluding the investigated one. For the Γ conduction band valley, it usually suffices to consider only the coupling to the highest valence bands. This leads to the well known Roth formula [74]

$$\mathrm{g}_c = 2 - \frac{2E_p\Delta_0}{3E_g\left(E_g + \Delta_0\right)}, \tag{4.2}$$

where the interband matrix element E_p describes the coupling strength, the band gap E_g gives the energetic distance to the valence bands, and the spin-orbit coupling energy is characterized by the parameter Δ_0. When the interband coupling is sufficiently strong and $\Delta_0 > 0$, the resulting g factor becomes negative. E.g. in bulk InAs one has $\mathrm{g}_c = -14.8$ [24].

The highest valence bands, on the other hand, are p-type bands that couple directly to the spin. Their bulk band-edge g factors can be derived from the six-band **k·p** Hamiltonian (1.17) as will be shown in the following. First, we replace the kinetic momentum $\mathbf{k}$ with the canonic momentum $\mathbf{K} = \mathbf{k} + (e/\hbar)\,\mathbf{A}$ in the 3×3 submatrices $\hat{H}^{3\times3}(\mathbf{k})$ of Eq. (1.18). Then, by utilizing the commutator relation

$$[K_i, K_j] = K_i K_j - K_j K_i = -\mathrm{i}\frac{e}{\hbar}\varepsilon_{ijk} B_k, \tag{4.3}$$

and by considering the **k·p** parameters defined in Eq. (1.20), we obtain the magnetic-field Hamiltonian [12]

$$\hat{H}^{3\times3}(\mathbf{K}) = \begin{pmatrix} LK_1^2 + M\sum\limits_{i=2,3} K_i^2 & N\{K_1, K_2\} & N\{K_1, K_3\} \\ N\{K_1, K_2\} & LK_2^2 + M\sum\limits_{i=1,3} K_i^2 & N\{K_2, K_3\} \\ N\{K_1, K_3\} & N\{K_2, K_3\} & LK_3^2 + M\sum\limits_{i=1,2} K_i^2 \end{pmatrix}$$

$$+ E_v + \frac{\hbar^2}{2m_0}\sum_{i=1}^{3} K_i^2 - \mu_B (3\kappa + 1)\sum_{i=1}^{3} \hat{I}_i B_i, \tag{4.4}$$

where we have used the abbreviations $N = N_+ + N_- = -6\gamma\hbar^2/(2m_0)$ and $\{K_i, K_j\} = (K_i K_j + K_j K_i)/2$. The last term in Eq. (4.4) results from the commutator [Eq. (4.3)] and gives the angular momentum contribution to the Zeeman splitting

$$\hat{H}^{6\times6}_{\mathbf{L}\cdot\mathbf{B}} = -\frac{\mu_B}{2}(6\kappa + 2)\,\hat{\mathbf{L}}^{6\times6}\mathbf{B}, \tag{4.5}$$

with $\hat{L}_i^{6\times6} = \hat{I}_i^{3\times3} \otimes 1^{2\times2}$ and the $L = 1$ angular momentum matrices $\hat{I}_i^{3\times3}$

$$\hat{I}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix}, \hat{I}_2 = \begin{pmatrix} 0 & 0 & \mathrm{i} \\ 0 & 0 & 0 \\ -\mathrm{i} & 0 & 0 \end{pmatrix}, \hat{I}_3 = \begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.6}$$

introduced in Ref. [12]. In addition, the coupling to the spin is given by the Pauli term from Eq. (1.34) (reduced to the six-band basis)

$$\hat{H}^{6\times6}_{\boldsymbol{\sigma}\cdot\mathbf{B}} = \frac{\mu_B}{2}\mathrm{g}_0\hat{\mathbf{S}}^{6\times6}\mathbf{B}, \tag{4.7}$$

where $\mathrm{g}_0 = 2$, and the matrix elements of the spin matrices $\hat{S}_i^{6\times6} = 1^{3\times3} \otimes \hat{\sigma}_i$ are given in terms of the Pauli matrices $\hat{\sigma}_i$. In the basis of the total angular momentum eigenstates $|J, J_3\rangle$ [Eqs. (1.22)-(1.24)], the total Zeeman term of the valence bands is given by [89, 90]

$$\hat{H}^{6\times6}_{\mathbf{J}\cdot\mathbf{B}} = \hat{H}^{6\times6}_{\mathbf{L}\cdot\mathbf{B}} + \hat{H}^{6\times6}_{\boldsymbol{\sigma}\cdot\mathbf{B}} = \frac{\mu_B}{2}\begin{pmatrix} -4\kappa\hat{\mathbf{J}} & 6(\kappa+1)\,\hat{\mathbf{U}} \\ 6(\kappa+1)\,\hat{\mathbf{U}}^\dagger & -(4\kappa+2)\,\hat{\sigma} \end{pmatrix}\mathbf{B}, \tag{4.8}$$

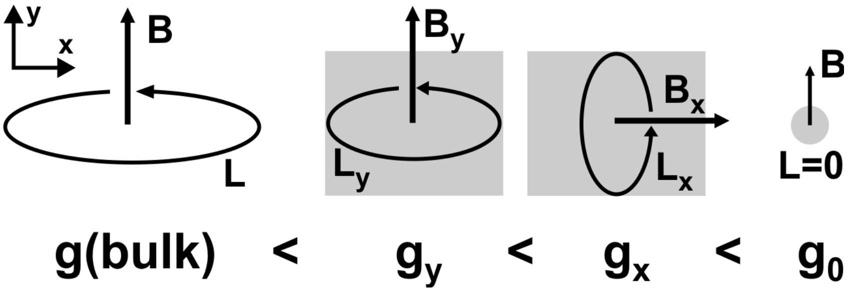


Figure 4.2: Schematic illustration of angular momentum quenching in a nanostructure (gray area) by the reduction of the dimension perpendicular to the magnetic field. The g factor scales with this size quantization.

with the 4 × 4 angular momentum $J = 3/2$ matrices $\hat{J}_i$ [12]. The 4 × 2 cross-space matrices $\hat{U}_i$ have been introduced in Ref. [89] and corrected in Ref. [90]. The Hamiltonian (4.8) leads to the anisotropic bulk hole g factors [66, 75, 78]

$$\mathrm{g}_{\|}^{hh} = -6\kappa,\ \mathrm{g}_{\perp}^{hh} = 0,\ \mathrm{g}_{\|}^{lh} = -2\kappa,\ \mathrm{g}_{\perp}^{lh} = -4\kappa, \tag{4.9}$$

where $\|$ and $\perp$ denote the directions relative to the quantization axis. The hole g factors are usually strongly negative, e.g., in InAs one has $\kappa = 7.68$ and therefore $\mathrm{g}_{\|}^{hh} \approx -46$ [24].

In nanostructures such as quantum dots, the determination of g factors gets significant more complicated. Here, confinement energy and strain may alter the energetic distances between states and therefore affect the spin-orbit coupling. The states are in fact mixtures from different bands and especially the valence band mixing strongly depends on the concrete geometry. In addition, wave functions can penetrate material interfaces and simultaneously occupy areas of different bulk g factors. Importantly, the orbital angular momentum of confined eigenstates is strongly affected by the geometry of the nanostructure. Besides possible spherical symmetry breaking, the envelope angular momentum gets quenched by any reduction of the available space in the plane perpendicular to the magnetic field [72]. As has been sketched in Fig. 4.2, this quenching depends on the orientation of the magnetic field with respect to the nanostructure, which in turn affects the g factors since the angular momentum is the source for the deviation of the g factors from 2. For electrons, the orientation dependence leads to an anisotropic g tensor rather than an isotropic g factor as it is the case in bulk [Eq. (4.2)]. Altogether, the g factors of semiconductor nanostructures lie somewhere between the bulk and the free-electron ones, but a quantitative analysis requires sophisticated theoretical models.

The calculation of g factors of realistic nanostructures requires a solution of the Schrödinger equation for a mesoscopic system. Since the g factors strongly depend on the spin-orbit interaction and the coupling between different bands, it is inevitable to use relativistic multiband models such as the eight-band **k**·**p** method. The spin splitting $\Delta E_{\uparrow\downarrow} = E^{\uparrow} - E^{\downarrow}$ of an electron or hole state is then determined from the eigenenergies of the spin-up and the spin-down solutions of the Schrödinger equation [Eq. (3.11)]. Since the spin splitting may depend on the direction of the magnetic field, we need to introduce a g tensor. So, for energies close to the electron and hole band edges, the magnetic-field dependent part of the Hamiltonian can be written as

$$\hat{H}_n = \frac{\mu_B}{2}\boldsymbol{\sigma} \cdot \hat{\mathrm{g}}_n \cdot \mathbf{B}, \tag{4.10}$$

where $\hat{\mathrm{g}}_n$ $(n = e, h)$ define the g tensors for the confined electron and hole states, respectively. The eigenvalues of this tensor yield the g factors $\mathrm{g}_n = \left(E_n^{\uparrow} - E_n^{\downarrow}\right) / \left(\mu_B B\right)$. We note that unless otherwise stated, we will consider only situations where the lateral confinement causes the electronic states to be well separated from higher lying states (cf. Fig. 4.1), so that this definition of g factors is unambiguous.

4.3 Nanowire quantum dots

4.3.1 Electron g tensor

In order to assess the accuracy of the present method, we calculate the g tensor of electrons in nanowire-based quantum dots. Experimentally, effective electron g factors of InP/InAs nanowire dots have been determined recently [86]. The quantum dots possess the wurtzite crystal structure, have a hexagonal shape, and are characterized by a diameter D, the InAs quantum dot length L, and a left and right InP barrier thickness w (see Fig. 4.3). The diameter D equals 50 nm in these experiments. The experimental values of the g factors range from $|\mathrm{g}| = 13$ which is close to the InAs bulk value for dots with $L = 270$ nm to $|\mathrm{g}| = 2.3$ for the thinnest dots with $L = 8$ nm.

In our calculations, we have simplified the geometry by assuming a quadratic or circular cross section of the dots (see inset of Fig. 4.3), since the precise shape of the dots does not influence our results in the range of the studied dot dimensions and magnetic fields. The **k·p** parameters for InAs and InP have been taken for the zincblende structure from Ref. [22] except for the κ-parameter that has been tabulated in Ref. [24]. The spin splitting $\Delta E_{\uparrow\downarrow} = E^{\uparrow} - E^{\downarrow}$ of the electron ground state in the dot has been calculated within a relativistic eight-band **k·p** model that is detailed in Sec. 1.2.1. The squared g tensor elements $G^{kl}, (k, l = 1, 2, 3)$ can then be extracted from the relation

$$\Delta E_{\uparrow\downarrow} = \mu_B \sqrt{\sum_{kl} B_k G^{kl} B_l}, \tag{4.11}$$

where B_i are the Cartesian components of the magnetic field. The spin splitting must be evaluated for a sufficiently small value of $|B|$ (which we took to be equal to 0.1 T) to exclude higher order contributions. Due to symmetry, only fields along the [100] and [001] directions need to be considered in the present case for solving the linear system of equations that result from Eq. (4.11). For our geometry and within the eight-band **k·p** model , we have $G^{xx} = G^{yy}$ and $G^{ij} = 0$ for all off-diagonal tensor components. We have determined the signs of the g tensor eigenvalues from the spin directions of the Zeeman split states and found that all calculated g factors had negative sign ($E^{\uparrow} < E^{\downarrow}$) as expected from the bulk values.

Figure 4.3 compares the calculated perpendicular effective electron g factor modulus $\mathrm{g}_{\perp} = (G^{xx})^{1/2}$ with the experimental results of Ref. [86] for dot lengths L between 8 and 20 nm and a barrier thickness of $w = 6$ nm. Furthermore, our calculations predict the effective electron g factor for the magnetic field parallel to the wire axis, $\mathrm{g}_{\|} = (G^{zz})^{1/2}$. The theory takes into account the strain as we will discuss in more detail below; Fig. 4.3 shows both types of results where strain effects in the Hamiltonian have been included or (artificially) set to zero to elucidate the sole strain effect on the g factor. As one can deduce from the figure, our calculations excellently reproduce the experimentally observed trend that shows the g factors to increase with increasing dot length. In semiconductors, g factors of conduction band electrons are generally dominated by the contribution from the angular motion. The more extended the electron wave functions are, the closer the g values lie to the InAs bulk value. The opposite limit of strongly confined electrons gives the pure spin value g = 2. This explains the trends in both the experimental and theoretical results. Since the electron wave functions are only weakly confined for magnetic fields along the wire, one expects the parallel g factors $\mathrm{g}_{\|}$ to be larger than the perpendicular values $\mathrm{g}_{\perp}$, and our calculations confirm this conjecture.

We note that electron g factors for such quantum dots have been calculated previously [91]. However, we found the hole g factors in Fig. 2 of the related Ref. [72] to disagree with the present method. We can quantitatively reproduce the results of this paper by using an incorrectly symmetrized ordering that can be obtained by substituting Eq. (4.12) into Eq. (1.18). We have checked that the use of the correct operator ordering in Eq. (1.18) plays a negligible role for the electron g factors in the nanowire dots. Thus, our results are consistent with Ref. [91].

4.3.2 Hole g tensor

We now turn to the hole g factors in these nanowire based quantum dots. For quantum wells in a magnetic field, each valence subband splits into a complicated pattern of Landau levels that are conventionally labeled by the orbital momentum J_z and an integer $n = -2, -1, 0, ...$ [92]. Since all those states with the same modulus $|J_z|$ become degenerate for zero magnetic field, these states tend to cross at some magnetic field which makes it nontrivial to uniquely define hole g factors. The situation in quantum dots is far more transparent. Here, the lateral confinement splits the states corresponding to different lateral momenta of the envelope function already at $B = 0$ (shown in Fig. 4.1 for an analogous situation). This allows one to uniquely define the hole g tensor by the energy separation between the energetically lowest dot state and

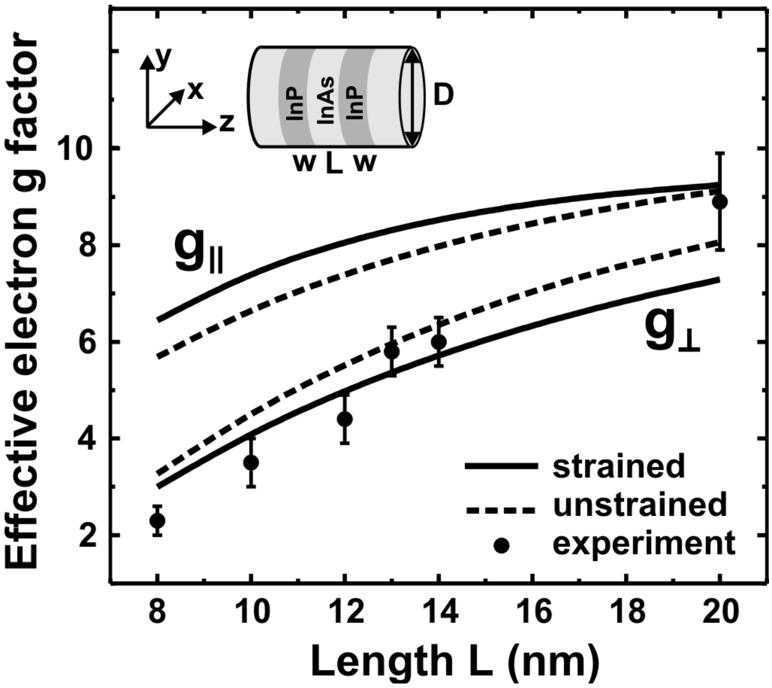


Figure 4.3: Comparison of calculated strained (solid lines), calculated unstrained (dashed lines) and experimental (circles) effective electron g factors for wire dots of length L. The $g_{||}$ and $g_\perp$ components correspond to the magnetic field lying parallel and perpendicular to the wire axis, respectively. Experimentally, only the $g_\perp$ components have been determined (Ref. [86]).

the lowest excited state with reversed sign of J_z. For the dots considered here, this pair of states is dominantly heavy-hole-like and originates from the pair of quantum well Landau levels with $J_z = -3/2$, $n = -2$ and $J_z = 3/2$, $n = 1$ [92]. These states remain the energetically lowest dot states up to magnetic fields of 4 T, and ensure an unambiguous definition of the hole g factor.

In Fig. 4.4, we show our calculated results for different dot lengths, both with and without the inclusion of strain, and for parallel and perpendicular magnetic fields. Again, all g factors are negative, but for simplicity we only discuss their absolute values. The size dependence of hole g factors in quantum dots is more complicated than for electrons. The dots considered here have a large width to height ratio so that carriers are only weakly confined laterally, i.e., perpendicular to the wire axis. Therefore, the calculated hole g factors in Fig. 4.4 resemble the values found for wide quantum wells, which can be obtained from the bulk Hamiltonian [12]. In the latter case, already a six-band $\mathbf{k}\cdot\mathbf{p}$ model gives the well-known results [Eq. (4.9)] $|g_{||}^{hh}| = 6\kappa$, $|g_{||}^{lh}| = 2\kappa$, $|g_\perp^{hh}| = 0$, $|g_\perp^{lh}| = 4\kappa$, where κ is a Luttinger parameter. For InAs, one has

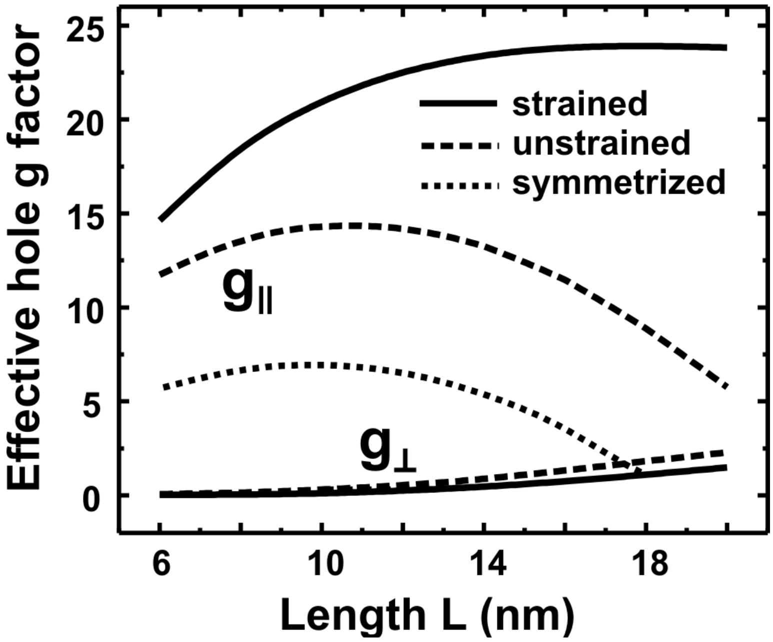


Figure 4.4: Comparison of calculated strained (solid lines) and unstrained (dashed lines) effective hole g factors for wire dots of length L and magnetic fields pointing perpendicular ($g_\perp$) and parallel ($g_{||}$) to the wire axis. The dotted line illustrates the importance of the correct operator ordering in Eq. (1.18) by showing the unstrained $g_{||}$ for an Hermitian but naively and therefore incorrectly symmetrized Hamiltonian.

$\kappa = 7.68$ which is much larger than this constant in InP, where one has $\kappa = 0.97$ [24]. Due to the confinement, the wire dot ground state is mainly heavy-hole-like and its g factors $g_{||}, g_\perp$ show an anisotropy that is somewhat smaller but still of the order of the bulk anisotropy. For the shortest dots, the wave functions penetrate into the InP barriers. This effect leads to a reduction in the g factors because κ is smaller for InP. Furthermore, the orbital motion is hampered in the limit $L \to 0$ so that the hole g factors tend toward the pure spin value in this limit. The larger the dot length L and the more extended the hole wave functions are, on the other hand, the closer the g factors lie to the bulk values of InAs. However, there is a competing effect that turns out to dominate in the limit of large values of L. The reduction in axial confinement for large L implies an increase in the light-hole (lh) contribution of the ground state. Indeed, the light hole contribution amounts to only 1.5% for $L = 8$ nm but increases to 12.0% for $L = 20$ nm. Since $|g_\perp^{lh}| > |g_\perp^{hh}|$ in bulk, $g_\perp$ increases also in dots in the limit of large L. Analogously, $g_{||}$ decreases in this limit since $|g_{||}^{lh}| < |g_{||}^{hh}|$ in bulk. Qualitatively similar trends have been predicted previously for self-assembled Si/Ge quantum dots [66], and for InGaAs dots [67, 85]. Experimentally, heavy-hole (hh) g factors have been determined for GaAs quantum wells [78, 79, 81], and for self-assembled InGaAs dots [83, 84]. In both of these situations, the observed values are much smaller than for InAs nanowire dots, since the bulk g factors of GaAs are much smaller ($\kappa = 1.2$), and carriers are significantly more strongly confined in self-assembled dots.

Finally, we have investigated the influence of the momentum operator ordering in the Hamiltonian of Eq. (1.18) on hole g factors. As has been discussed previously (cf. Sections 4.2.2 and 1.2.1), the known continuum limit of the **k·p** Hamiltonian [12] imposes an unambiguous constraint on the ordering of momentum operators in the discrete Hamiltonian [17, 19]. To investigate the effect of this operator ordering, we have replaced the correct Hamiltonian in Eq. (1.18) by an incorrect one, namely a naively symmetrized version. To this end we set [see Eq. (1.18)]

$$N'_{+} = N_{-} = \left(N'_{+} + N_{-}\right)/2. \tag{4.12}$$

The resulting hole $g_{\|}$ factors for quantum wire dots are included in Fig. 4.4 (dotted line) and are seen to deviate strongly from the correctly calculated values.

4.3.3 Strain effects

The InP/InAs nanowires investigated by Björk et al. [86] are free standing wires with a relaxed InAs lattice structure and pseudomorphically strained InP layers. We have calculated the spatial strain profile of the heterostructure wires by minimizing the total elastic energy in a linear continuum elasticity model [39] (cf. Sec. 1.4). Fig. 4.5 (a) shows the resulting strain profile along the wire axis near the dot with a length of $L = 20$ nm and barrier thicknesses of $w = 6$ nm. We find the InAs dot to be compressively strained ($\varepsilon_{xx} = \varepsilon_{yy} < 0$), whereas the InP barriers show a tensile strain. Outside of the InP barrier material, the lattice relaxes to unstrained InAs after approximately 25 nm. Our calculations are qualitatively consistent with previous experimental and theoretical strain results obtained for similar wires with much smaller diameters and a single InAs/InP interface [93]. The incorporation of this strain into the Hamiltonian via linear deformation potentials (cf. Sec. 1.4.2) leads to changes in the electronic structure and the g factors. For the electron g factor, these trends can already be understood in a bulk model originally derived for electrons in homogeneously strained layers [94]. The band gap within the dot is slightly increased due to the hydrostatic pressure component of the strain, $\delta E_g = (a_c - a_v)(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}) > 0$, and the heavy-hole band edge gets pushed above the light-hole band edge due to a tetragonal distortion that is proportional to $\varepsilon_{zz} - \varepsilon_{xx}$ [see Fig. 4.5 (b)]. This distortion additionally induces a mixing of the light-hole and the split-off (so) hole band. Together, these changes in the electronic structure affect the coupling between conduction and valence bands and enhance the anisotropy of the electron g factor [94]. Namely, the increased band gap reduces the conduction band - valence band coupling, which is the source for the deviation of the electron g factors from 2. For the perpendicular g factor, the lh/so mixing adds up to this effect. Therefore $g_{\perp}$ is decreased, still reproducing the experimental results. By contrast, for

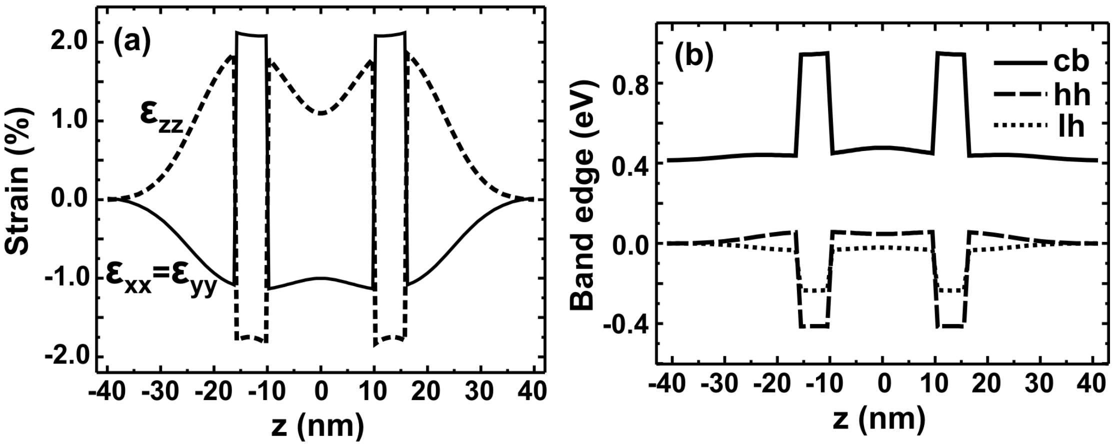


Figure 4.5: (a) Cross section of calculated strain components along the wire axis z for a dot with a length of $L = 20$ nm and a barrier thickness of $w = 6$ nm. (b) Calculated energies of conduction (cb), heavy hole (hh), and light hole (lh) valence band edges.

$g_{||}$, the lh/so mixing has an increasing effect. Here the hh/lh splitting additionally raises the g factor. Together, these effects exceed the general reduction due to the increased band gap and lead to larger values of $g_{||}$.

Hole g factors are far more sensitive to strain effects because the hole states are p-states and couple directly to the spin. The main effect of the compressive strain within the dot is to shift the heavy-hole band edge above the light-hole band edge. This reduces the light-hole contribution to the ground state from 1.5% in the unstrained case to 1.1% for $L = 8$ nm and from 12% to 3.8% for $L = 20$ nm. Again following the trends in the bulk values, this implies a slight increase in $g_{\perp}$ and a decrease in $g_{||}$ that is much less pronounced than in the unstrained case. This explains the increasing difference between the strained and unstrained g factors in the limit of large dot lengths L.

4.3.4 Spin splitting for high magnetic fields

Since our approach takes into account magnetic fields nonperturbatively, we may study the electronic structure at high magnetic fields B. In the framework of $\mathbf{k}\cdot\mathbf{p}$ theory, however, the magnetic length must remain large compared to the lattice constant (i.e. $B \ll 1000$ T). In the magnetic field regime above 4 T, the Zeeman splittings can exceed the energy differences between ground and excited states of the dot (similar to Fig. 4.1). Therefore, the energetic order of alternating spin-up and spin-down states gets mixed up. However, the order of ground and excited states with the same spin direction remains, and the states that account for the ground state g factor do not alter. In addition, the strong confinement and strain induced light- and

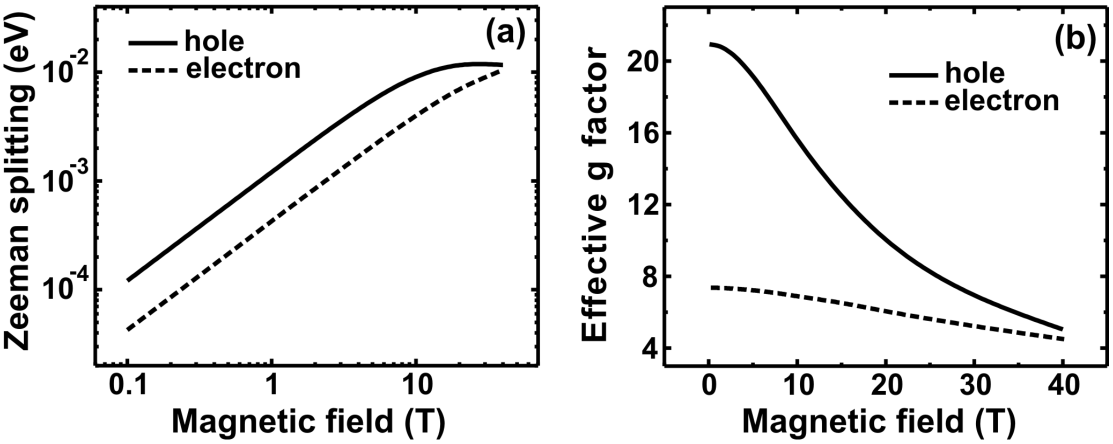


Figure 4.6: (a) Calculated electron (dashed lines) and hole (solid lines) Zeeman splittings and (b) effective g factors for electrons and holes as a function of the magnetic field lying parallel to the wire axis for a dot of length $L = 10$ nm.

heavy-hole splittings guarantee that all light-hole states remain energetically well separated from the heavy-hole ground state up to the highest B fields considered, so that the hole g factor remains unambiguously defined. The spin splitting of the lowest electron state increases linearly with B up to 10 T for a dot of length $L = 10$ nm and the magnetic field pointing along the wire axis. The spin splitting of the hole ground state responds more sensitively to B and deviates from linearity already at 4 T. In Fig. 4.6, we depict the electron and hole $g_{||}$ factor up to 40 T. We see that the effective $g_{||}$ factors $\Delta E_{\uparrow\downarrow}/(\mu_B B)$ decrease with increasing field strength, which resembles the experimentally observed trends in quantum wells [79, 80]. We associate this effect with a magnetic-field induced band mixing. For the hole ground state, we find the average heavy-hole component of the spin-up and spin-down states to decrease with increasing B. This diminishes the $g_{||}$ factor of holes as has been pointed out previously [67]. We find an analogous but less pronounced behavior for electrons, where the conduction-band component decreases with increasing B.

To visualize the influence of high magnetic fields, in Fig. 4.7, we show the lateral deformation of the electron ground state wave function within an $L = 20$ nm dot as a function of a magnetic field $B_\perp$ perpendicular to the wire axis. This wave function deformation is controlled by the ratio between the magnetic cyclotron energy $\hbar\omega_c$ and the lowest zero field electron excitation energy ΔE in the dot. The latter quantity is a measure of the confinement. The larger the value of ΔE, the smaller the deformation induced by the magnetic field is [72] and this can be clearly deduced from the figure.

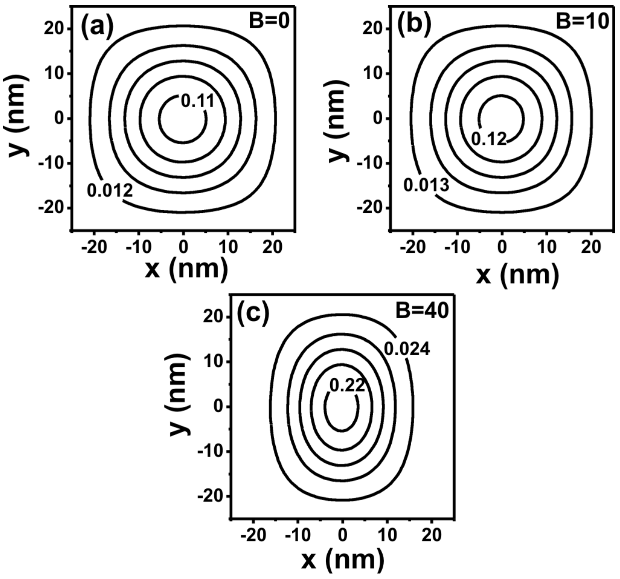


Figure 4.7: Cross section of the density of the electron ground state at the dot center in units of 10^{18} cm^{-3} for different magnetic fields in Tesla. The ground state's density shows a noticeable deformation, if the ratio $\alpha = \hbar\omega_c/\Delta E$ is close to one. (a) $\alpha = 0$, (b) $\alpha = 0.2$ (c) $\alpha = 0.8$.

4.4 Self-assembled quantum dots

4.4.1 Size variation of g tensors

In this section, we investigate self-assembled InGaAs dots that are embedded in GaAs and are subject to external magnetic fields. Although the g factors of self-assembled dots show less variation than the nanowire dots from the last section, they are of particular interest since the experimental realization of these dots is much better established. In addition, the g factors can have different signs so that we now need to consider the full g factors, rather than only the effective moduli as in the previous sections. The dots considered here, are characterized by the quantum dot height h, the dot width w, and the alloy profile within the InGaAs dot [see inset in Fig. 4.8(a)] that depends on the growth conditions. We investigate dot heights h from 3 to 7 nm, widths w from 15 to 25 nm, and use a trumped-shaped alloy profile throughout. This alloy profile is characterized by an indium distribution that starts from 80% at the tip and decreases to 30% toward the bottom corners [43]. We model the dots as truncated pyramids that sit on 1 nm $\mathrm{In_{0.5}Ga_{0.5}As}$ wetting layers on a (001)-GaAs substrate. By comparison of calculated and experimental Stark shifts [7, 95], these structural parameters have been shown to be realistic for dots with a nominal indium content of 50%. For the present structure, we find the principal axes of all g tensors [Eq. (4.10)] to be given by the vectors [001], [110], and $\left[1\bar{1}0\right]$.

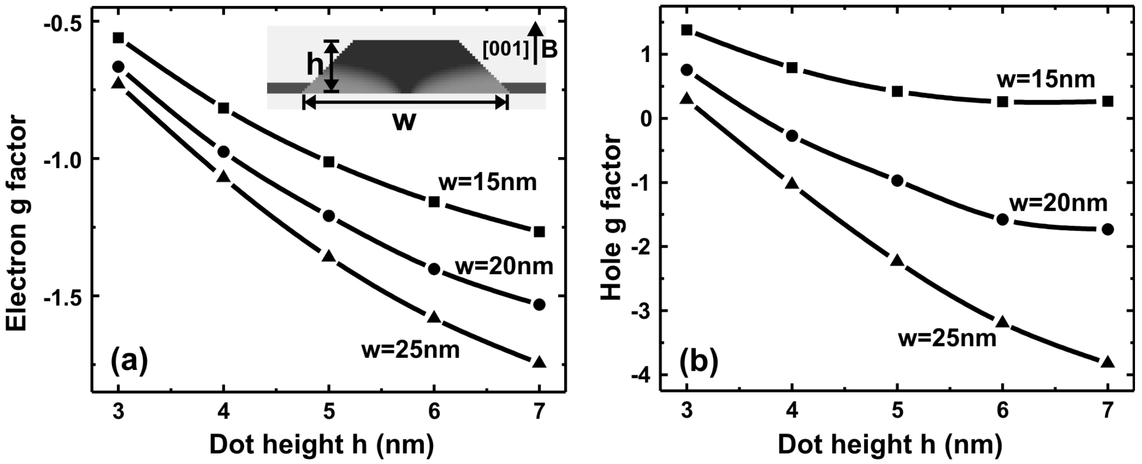


Figure 4.8: (a) Calculated electron ground state g factors of self-assembled InGaAs quantum dots as a function of the dot height for different dot widths. The magnetic field lies in the vertical [001] direction. The inset shows a schematic cross section of the structure studied. (b) Same for hole ground state.

In Fig. 4.8(a), we show the g factor associated with the electron ground state in the dot for the magnetic field lying in the [001] direction. Compared to the InAs nanowire-based dots of Sec. 4.3.1, the electron g factor shows very similar but less pronounced trends. The larger the dot, the more extended is the electron wave function and the closer the g factor lies to the large negative bulk value. However, due to the strong confinement in self-assembled dots, the overall range of g factors lies much closer to the pure spin limit of g = 2 than for the InAs wire-dots. For the hole ground state g factor shown in Fig. 4.8(b), we obtain a much stronger variation with the dot width. In addition, we find the dot height to matter only for wide dots. For the smallest dots ($w = 15$ nm), we associate the increasing height to width ratio to be responsible for the almost constant g factor. Similar to the trends described in Sec. 4.3.2, the light-hole contribution to the dominantly heavy-hole ground state increases with the dot height. This leads to a decrease in the g factor modulus that partly cancels the overall increase when moving toward the bulk limit.

The two remaining principal axes of the g tensor are given by the diagonals [110] and $\left[1\bar{1}0\right]$ of the pyramidal base plane. Fig. 4.9(a) compares all g tensor eigenvalues of the electron and hole ground states for a fixed dot width $w = 25$ nm and heights between 3 and 5 nm. We find the g factors to differ strongly between the vertical [001] direction and the growth plane $[110] \times \left[1\bar{1}0\right]$. Within this plane, however, the anisotropy is almost negligible. These findings have also been observed experimentally [84, 87]. The calculated in-plane hole g factors are close to zero similar to the heavy-hole value in bulk [Eq. (4.9)]. For electrons, the g factor anisotropy is less pronounced. Here, the in-plane g factors have a smaller absolute value since the confinement of

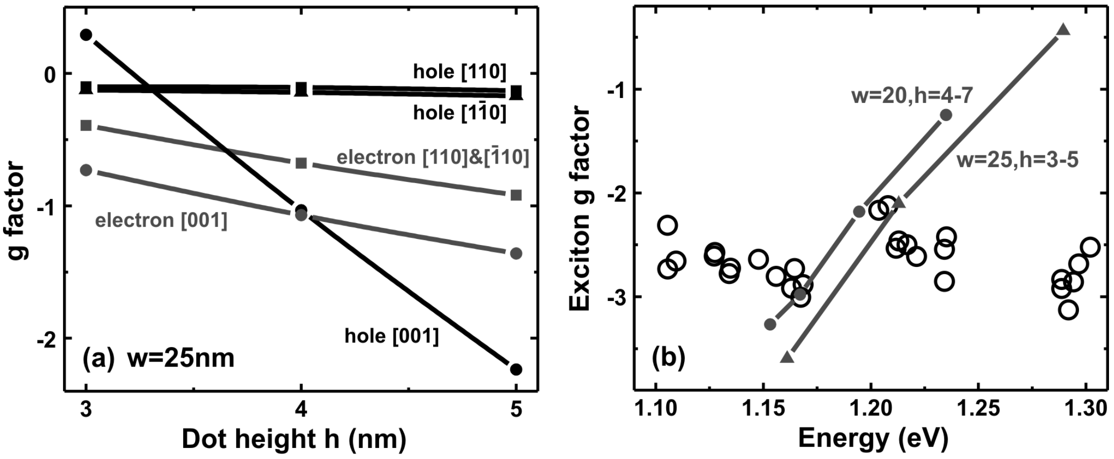


Figure 4.9: (a) Calculated electron and hole ground state g tensor elements along the principal axes for a dot width of $w = 25$ nm, as a function of the dot height. (b) Calculated exciton g factors (solid lines), as a function of the dot size parametrized by the transition energy between electron and hole ground state. The magnetic field lies in the [001] direction. Experimental values (circles) are taken from Ref. [85].

the electron wave function in the vertical direction is stronger than in the lateral directions. All g factor values are in reasonably well agreement with the experimental results from Ref. [84]

$$g^e_{[001]} = -0.80,\ g^e_{[110]/[1\bar{1}0]} = -0.65,\ g^h_{[001]} = -2.20,\ g^h_{[110]/[1\bar{1}0]} = -0.35. \tag{4.13}$$

A more detailed comparison with experiment is complicated by the fact that structural parameters of self-assembled quantum dots are rarely known exactly and seem to vary strongly. In addition, experimental determination of electron and hole g factors is rather difficult. Much more data exists for exciton g factors that are basically given by the sum of electron and hole g factors $g_x = g_e + g_h$ [83] (see also Sec. 5.2). In Fig. 4.9(b), we show calculated exciton g factors for the magnetic field lying in the [001] direction, as a function of the dot size, parametrized by the fundamental energy transition between electron and hole ground state $E(e_0) - E(h_0)$. Comparison with experimental data from Ref. [85] shows reasonable agreement and demonstrates the wide distribution of exciton energies for nominally equal dots.

4.4.2 Electric tuning of g factors in single dots

Since electric control of g factors would be highly desirable, we finally investigate the influence of external electric fields on the g factors of single self-assembled quantum dots. In Fig. 4.10(a), we show the g factors of the electron and hole ground state, when the dot is subject to an electric field as well as a magnetic field, both lying

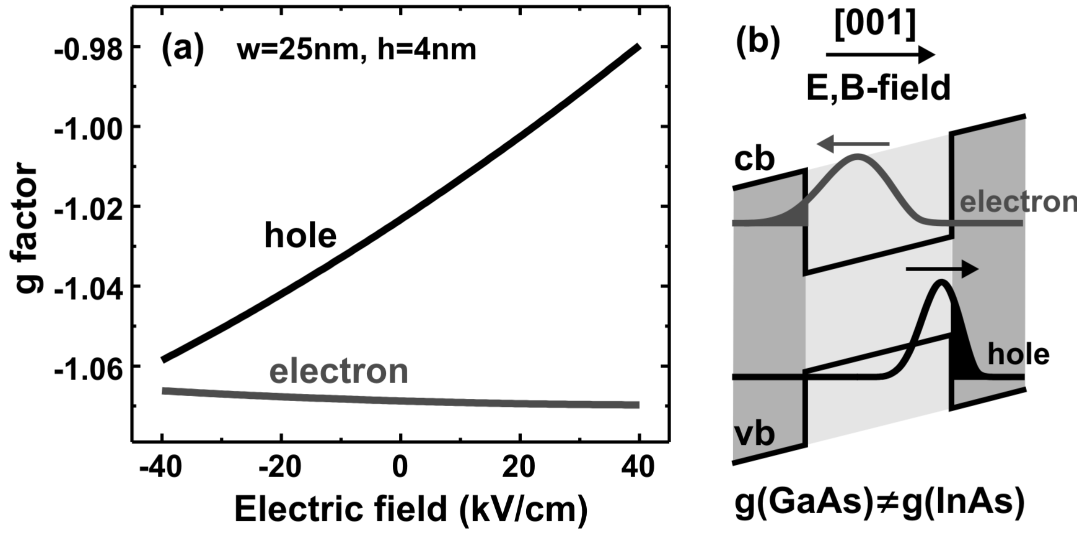


Figure 4.10: (a) Calculated electron and hole g factors for a dot size of $w = 25$ nm and $h = 4$ nm, as a function of the vertical electric field. The magnetic field lies in the [001] direction, aswell. (b) Schematic cross section of band edges and probability densites along the vertical direction. The figure illustrates the stronger localization of hole states that leads to a better tunability of hole g factors.

in the vertical [001] direction. The electric field has been introduced by applying a bias between a top and a back gate, as has been discussed in Sec. 1.3.2. For the dot size, we assume the realistic dimensions of $h = 4$ nm and $w = 25$ nm that lead to almost equal electron and hole g factors. The electric field shifts the positions of the electron and hole wave functions within the dot, as has been sketched in Fig. 4.10(b). This affects the distribution of the probability density over areas of different indium content and the overlap with the GaAs barrier that has a different bulk g factor. For the hole ground state, we predict an almost linear dependence of the g factor on the electric field strength. The tunability amounts to a few percent for reasonable electric field values. By contrast, the electric field dependence of the electron g factor is found to be negligible. This originates from the fact that the electron wave function is more delocalized than the hole, due to its smaller effective mass. Therefore, the electron perceives an average over a larger area of different chemical environments and its g factor is fairly insensitive to slight modifications of the probability distribution. In addition, the overall range of electron g factor modulation is smaller than for holes, as can be deduced from Fig. 4.8. In total, we have to conclude that the tunability of g factors in single quantum dots is rather limited and more complicated structures will be required to allow realistic applications.

4.5 Summary

In this chapter, we have investigated the electronic structure of quantum dots in high magnetic fields. From the magnetic energy spectrum, we have highlighted the gyromagnetic factor as one of the most intriguing and application relevant properties of semiconductor nanostructures in magnetic fields. Starting from an overview on the origin of g factor variations in bulk semiconductors and nanostructures, we have illustrated our method from chapter 3, by a quantitative analysis of g factors in quantum dots. Concretely, we have computed the g tensor for electrons and holes in InP/InAs nanowire-based quantum dots. For electrons, we have obtained excellent agreement with experimental data [86]. Our results show that the changes in the electron g factors correlate well with the spatial extent of the wave functions and therefore with the angular motion, as has been argued before [72]. The g factors for holes are shown to depend much more sensitively on the sample geometry and material composition which allows significant tailoring of their values, e.g., by changing the dot size. We predict a strong anisotropy of the g tensor components for magnetic fields that lie parallel or perpendicular to the wire axis. We find this anisotropy to get enhanced with strain. Additionally, we have calculated the influence of high magnetic fields on the g factors and find markedly nonlinear Zeeman splittings. Finally, we have extended our investigation to self-assembled InGaAs quantum dots. Here, we have calculated electron and hole g tensors as a function of the dot size and investigated their tunability in an electric field. In terms of the size dependence, we find similar but less pronounced trends than for the nanowire-based dots. The limited bias control over g factors in single quantum dots raises the question about more effective devices that will be discussed in the next chapter.

Chapter 5

Electrically controllable g tensors in quantum dot molecules

5.1 Introduction

Spins of confined carriers in quantum dots are promising candidates for the logical units in quantum computers [1, 96, 97]. In many concepts developed so far, the individual spin-qubits are being manipulated by magnetic fields [98], which is difficult to achieve in practice. An alternative procedure is to address individual spin-qubits by their own electric gate [99] which allows fast changes of the spin splitting, quantum gate operations and a tuning of the spin storage time [97]. In any case, progress in this field requires a detailed understanding of the mechanisms that allow one to modify the spin-related electronic structure properties such as gyromagnetic factors. In heterostructures, the possibility of electric control of g factors has been demonstrated by shifting the wave functions between different material regions by an applied bias [100]. Together with the anisotropy of the g tensor, such an electrically controlled tuning of the Zeeman splittings allows spin manipulations without time-dependent magnetic fields [101]. These experiments have been important proofs of principle but still represent ensemble averages. In quantum dots, electron and hole g factors have been extensively studied experimentally [85, 86] and theoretically [67, 72] (cf. Sec. 4.3). In self-assembled dots, the g factors have been found to be almost isotropic within the growth plane [84, 87] (cf. Sec. 4.4.1). In addition, the tunability of quantum dot g factors, which has been investigated both theoretically in Sec. 4.4.2 and experimentally [102] is rather limited, because the bound quantum dot wave functions are fairly insensitive to applied electric fields. By contrast, distinctive electrically tuned resonances have been discovered in recent experiments for the exciton g factors in vertically stacked quantum dot molecules [103]. It is plausible that coupled quantum dots provide more room for shifting the electron and hole wave functions between the two quantum dots by an external electric field. Indeed, the authors have been able to explain their results in terms of a simple phenomenological model based on the bonding and antibonding nature of the molecular states [103].

In this chapter, we significantly extend the theoretical analysis of quantum dot molecules by showing that the piezoelectric polarization associated with stacked quantum dots produces a giant electrically tunable *in-plane* anisotropy of hole g factors. This g factor modulation is an order of magnitude larger than in individual quantum dots and effectively allows g factor switching between almost zero and a constant finite value by a single electric gate. We provide a general and quantitative theoretical analysis of the electric field dependent electron, hole, and exciton g tensors of coupled quantum dots. For vertical magnetic fields, we show that our calculated results reproduce the experimentally observed exciton g factors in Ref. [103] without any fitting parameters.

The chapter is organized as follows. In Sec. 5.2, we discuss the method employed for our calculations of the electron, hole, and exciton g tensors in quantum dot molecules. Results for the bias driven tuning of g factors in vertical magnetic fields are given in Sec. 5.3.1. Here, we also compare our results with experimental data [103]. In Sec. 5.3.2, we show that g tensor components can be switched on and off by an electric field and study the tunability of the g tensor anisotropy in lateral magnetic fields. These results are used to propose a universal single spin-qubit gate in Sec. 5.3.3. Finally, in Sec. 5.3.4, resonant electron g factors in molecules with larger dot separation are studied.

5.2 Method and structure

The realistic calculation of band-edge g factors of a self-assembled quantum dot molecule requires a solution of the Schrödinger equation for a mesoscopic system. It must include the substrate, the wetting layers, as well as the overgrown quantum dots, and must take into account the macroscopic strain field, the piezoelectric polarization, as well as any applied magnetic and electric fields. Only recently it has been recognized that an accurate incorporation of the magnetic field into the Schrödinger equation requires special care to ensure gauge-invariant results [67, 72] (cf. chapter 3). Since a key point of this work lies in the tuning of g factors by an applied bias, both a magnetic and an electric field will be incorporated into the Schrödinger equation. We calculate the energies of electron and hole ground and first excited states of the entire mesoscopic system using our relativistic eight-band **k·p** envelope function method that has been described in detail in Sections 1.2.1, 3.3 and has been implemented into next**nano**++ [4]. The Hamiltonian can be written schematically in the form

$$\hat{H} = \hat{H}^{8\times 8}_{\mathbf{k}\cdot\mathbf{p}}(\mathbf{x}, \mathbf{x}', \mathbf{B}) + \frac{g_0 \mu_B}{2} \hat{\mathbf{S}}^{8\times 8} \cdot \mathbf{B} + e\,\mathbf{x} \cdot \mathbf{F}, \tag{5.1}$$

where the first term on the right-hand side represents the eight-band effective mass Hamiltonian of the entire structure in a discrete real-space basis (embracing N grid nodes). This term includes the coupling to the magnetic field $\mathbf{B}$ in a manifestly gauge-invariant manner with $\mathbf{B}$ only appearing in phase factors. Strain effects are incorporated into the Hamiltonian via linear band-edge deformation potentials (cf. Sec. 1.4.2) and piezoelectric charges (cf. Sec. 1.4.3). The three-dimensional strain field is calculated by globally minimizing the total elastic energy, employing a continuum elasticity model [39], as has been discussed in detail in Sec. 1.4. We assume the material interfaces to be governed by pseudomorphic growth. In addition, we assume vanishing normal stress at the boundaries of the simulation domain. For localized nanostructures embedded into a substrate, the size of this three-dimensional domain is increased until the lattice has relaxed to the unstrained substrate on all boundaries. For the dots considered here, this requires a domain size of the order of 100 nm along each axis. The second term on the right-hand side of Eq. (5.1) couples the spin to the field $\mathbf{B}$. Here, μ_B is the Bohr magneton, $\mathrm{g}_0 = 2$ is the free-electron g factor, and the 8×8 spin matrices $\hat{S}_i$ $(i \in \{x, y, z\})$ [Eq. (1.35)] are completely determined by the Pauli matrices. The homogeneous electric field $\mathbf{F}$ is assumed to point along the vertical [001] growth direction and has been applied via two contacts, as discussed in Sec. 1.3.2. The $8N$ dimensional Hamiltonian does not include free-carrier charges but is augmented by the Poisson equation [Eq. (1.65)]; in this way, the piezoelectric polarization is fully taken into account. We will focus on bound eigenstates of this Hamiltonian that are twofold Kramers degenerate at $\mathbf{B} = 0$ and do not get split by inversion asymmetry effects. This applies to all s-type envelope states such as quantum dot ground states but not to some higher lying excited states, as has been discussed in detail in Ref. [104]. For nonzero but small magnetic field, the states are subject to a Zeeman splitting that we find to depend linearly on the $\mathbf{B}$ field up to approximately 10 T for the present structures. For the g tensors $\hat{\mathrm{g}}_n$ $(n = e, h)$ of the confined electron (e) and hole (h) states, we use the definition of Eq. (4.10). The g factors g_n that represent the eigenvalues of these tensors can again be determined from the Zeeman splittings $\mathrm{g}_n = \left(E_n^{\uparrow} - E_n^{\downarrow}\right) / \left(\mu_B B\right)$. For both the electron and hole band edges, we focus on the ground (0) and first excited (1) state.

The g factor g_x of a neutral magnetoexciton X^0 is defined by the energy difference between the configurations $\sigma^+ \left(e^{\downarrow} h^{\uparrow}\right)$ and $\sigma^- \left(e^{\uparrow} h^{\downarrow}\right)$, where electron and hole states have opposite spins,

$$\mathrm{g}_x = \frac{E\left(\sigma^-\right) - E\left(\sigma^+\right)}{\mu_B B}. \tag{5.2}$$

We note that the other combinations of electron and hole states do not lead to optically active configurations. Since the electron-hole exchange energy is small compared to typical Zeeman splitting energies, we define the exciton g factors g_x^0 and g_x^1 for the lowest and first excited magnetoexciton state by [83]

$$\begin{aligned} \mathrm{g}_x^0 &= \mathrm{g}_e^0 + \mathrm{g}_h^0, \\ \mathrm{g}_x^1 &= \mathrm{g}_e^0 + \mathrm{g}_h^1, \end{aligned} \tag{5.3}$$

where g_e^0 and g_h^0 are the electron and hole ground state g factor, respectively. Since the hole states lie much denser than the electron states, the excitonic g factor g_x^1 involves the ground electron g_e^0 and first excited hole g factor g_h^1.

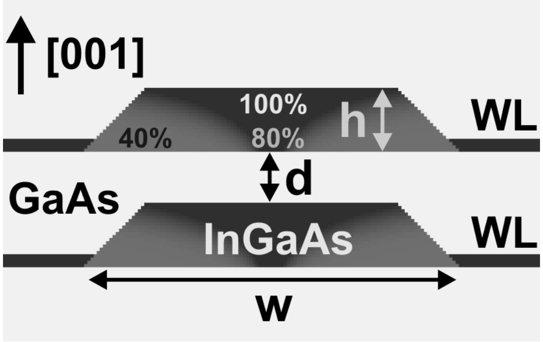


Figure 5.1: Schematic cross section of vertically stacked (001)-grown InAs/GaAs double dot structure studied in this chapter. We assume a height of $h = 2.5$ nm, a width of $w = 15$ nm, dot distances of $d = 1.5$ or 2 nm, a wetting layer (WL) thickness of 0.5 nm, and a realistic trumpet-shaped alloy profile within the dots (Ref. [107]).

For the concrete predictions presented in this chapter, we consider vertically stacked InAs/GaAs double dot structures that have been fabricated and studied experimentally recently [105, 106]. The quantum dot molecule itself is characterized by the dot separation d, the quantum dot height h, the dot width w, and the alloy profile within the individual dots that we assume to have identical size and composition (see Fig. 5.1). We take $h = 2.5$ nm, $w = 15$ nm, $d = 1.5$ or 2 nm, and a realistic [43, 95] trumped-shaped alloy profile throughout. This alloy profile is described in detail in Ref. [107] and is characterized by an indium distribution that starts from 100% at the tip, and decreases to 80% and 40% toward the bottom center and the bottom corners of the individual dots, respectively. We model the dots as truncated pyramids with $\{011\}$ side facets that sit on 0.5 nm InAs wetting layers on a (001)-GaAs substrate. With these parameters, we find the lowest interband transition energy of the entire structure to be 1.30 eV for $d = 1.5$ nm as well as $d = 2$ nm. For this structure, we predict the principal axes of all g tensors to be given by the set [001], [110], and $[1\bar{1}0]$.

5.3 Results

5.3.1 Resonant tuning of exciton g factors

We first discuss the molecular eigenstates for zero magnetic field as a function of the vertically applied electric field in terms of the individual electron and hole ground states. For dot distances $d \leq 2$ nm, the lowest molecular electron state ranges over both dots and forms an extended bonding state for all electric fields considered here. This is a consequence of the small electron mass and has been visualized in Fig. 5.2. By contrast, the individual dot hole states are more localized and therefore respond

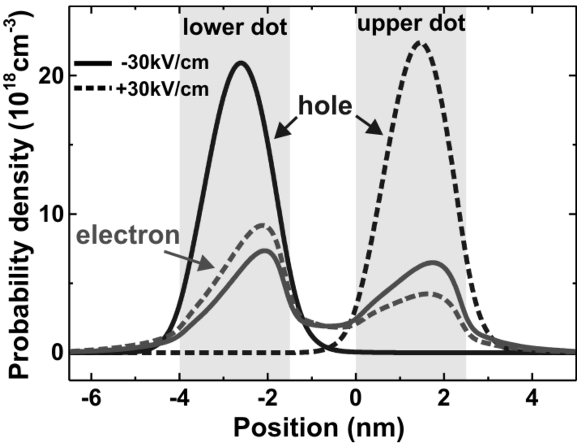


Figure 5.2: Cross sections of calculated molecular electron and hole ground state probability densities for a dot separation of $d = 1.5$ nm. The cross sections are taken at the dot center and run along the vertical [001] growth direction. In addition, we have applied a vertical electric field of +30 (solid lines) or −30 kV/cm (dashed lines), relative to the resonance field $\mathbf{F}_{\mathrm{res}}$.

more sensitively to the electric field. The field tunes the energies of these individual dot hole states relative to each other. For negative electric field ($\mathbf{F}\|[00\bar{1}]$), the molecular hole ground state is predominantly localized in the lower dot (cf. Fig. 5.2) and vice versa for positive electric fields. For a particular field $\mathbf{F}_{\mathrm{res}}$, the two ground states are tuned into resonance and form molecular bonding and antibonding orbitals with energies $E_{h,B}$ and $E_{h,A}$, respectively. This field strength $\mathbf{F}_{\mathrm{res}}$ is slightly nonzero because the strain field produces a small asymmetry between the individual dot energies [106]. For a dot separation of $d = 1.5$ nm, our calculations give $\mathbf{F}_{\mathrm{res}} = 7$ kV/cm. For this dot separation, we find $(E_{h,B} - E_{h,A})/2 = +0.6$ meV, which implies the bonding state to be the molecular hole ground state. By increasing the separation d between the dots to a value $d = 2$ nm, we find the role of bonding and antibonding states to become reversed and we obtain $(E_{h,B} - E_{h,A})/2 = -1.1$ meV. This is in good agreement with a previous theoretical result [108]. Experimentally, one finds a qualitatively similar trend but the crossover dot distance seems to lie at slightly larger values of the dot separation [103, 108].

We now turn to the molecular electron and hole g factors in the presence of a small vertical magnetic field $\mathbf{B}\|[001]$. In Fig. 5.3(a), we show the molecular hole g factors associated with the ground and first excited state, respectively. The hole g factors can be tuned by more than 100% by the electric field and show a pronounced resonance behavior associated with the formation of bonding and antibonding hole states. Qualitatively, this resonance formation has been explained previously in terms of a simple model [103]. Since, however, this work assumed incorrect signs of the bulk and single dot hole g factors, we present a brief discussion of our results. The heavy

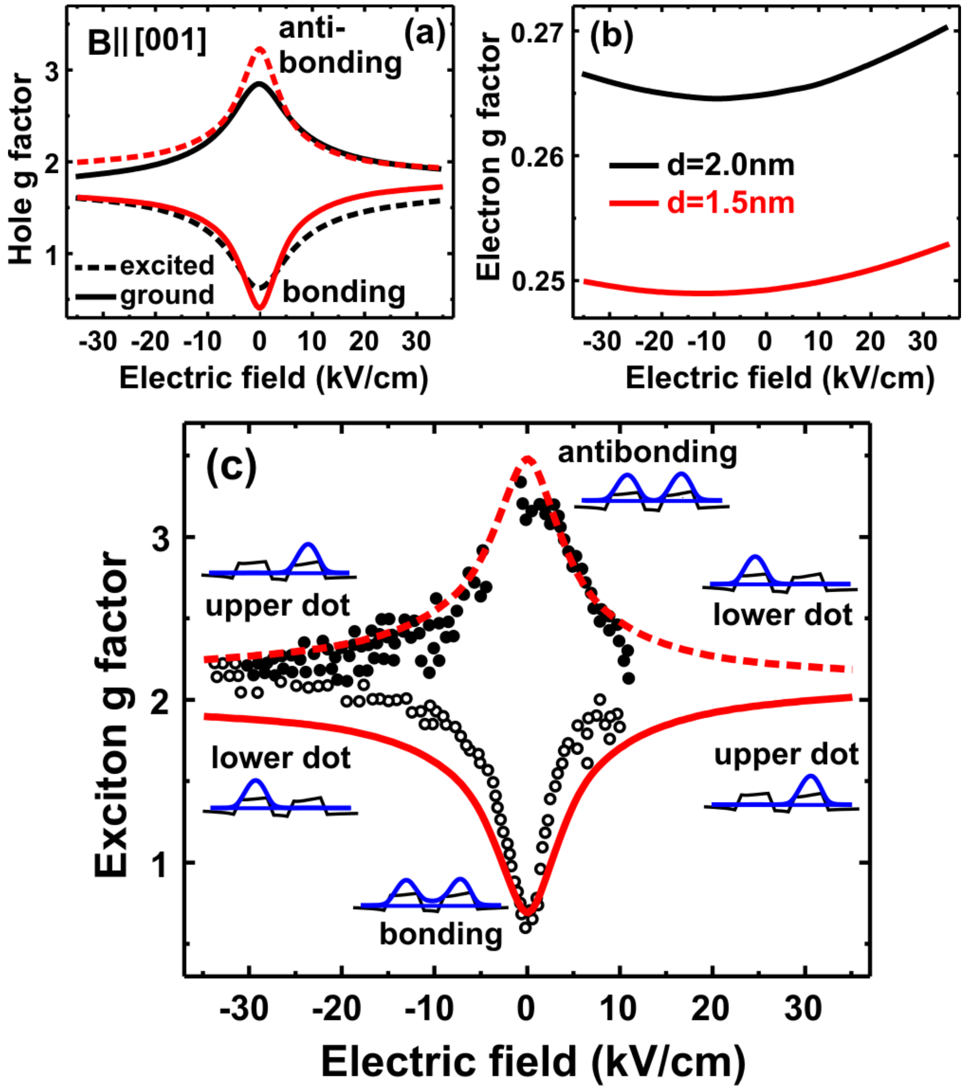


Figure 5.3: g factors of coupled quantum dots, for magnetic and electric fields lying in the vertical [001] direction. All electric field values are specified relative to the resonance field $\mathbf{F}_{\rm res}$, as discussed in the main text. (a) Calculated hole g factors for ground state (full lines) and excited state (dashed lines) for dot separations of $d = 2$ nm (black lines) and $d = 1.5$ nm (red lines), respectively. (b) Calculated electron g factors for the ground state. Dot separations are taken as in (a). (c) Comparison of calculated neutral exciton g factors with experimental results from Ref. [103] (circles). The excitons are formed by an electron in the ground state and a hole either in the ground (full line) or first excited state (dashed line). The dot separation is $d = 1.5$ nm. The insets indicate schematically the probability density of the hole states for different electric field values.

hole g factor in bulk GaAs is negative and given by $g_{hh}(GaAs) = -6\kappa = -7.2$ [Eq. (4.9)] in terms of the Luttinger parameter κ [24]. The present quantitative calculations show that the hole g factors associated with the isolated quantum dots are positive and approximately equal to +1.8. This result can be deduced from Fig. 5.3(a) in the limit of large positive or negative field where the molecular states are localized within the individual dots and the molecular coupling plays no role. We would like to point out that the present theory does predict negative hole g factors for larger quantum dots where quantum confinement is less pronounced (cf. Secs. 4.3, 4.4). This is also in agreement with experiment [83]. For the present coupled quantum dots, the formation of bonding (antibonding) states as a function of the electric field increases (decreases) the overlap of the molecular states with the GaAs barrier region in between the coupled quantum dots which leads to a decrease (increase) of the molecular hole g factor. The effect is more pronounced for smaller interdot distances and this is shown in Fig. 5.3(a) for the two cases of $d = 1.5$ and $d = 2$ nm. We note that the figure exhibits a slight asymmetry between positive and negative electric fields which is related to the different strain fields felt by the two dots [106]. The g factor associated with the electron ground state remains almost constant for the entire range of fields as shown in Fig. 5.3(b).

Based on these calculations of electron and hole g factors, we can now predict the excitonic molecular g factor. In Fig. 5.3(c), we show the resulting effective exciton g factors [Eq. (5.3)], formed by the electron ground state and the hole ground and first excited state as a function of the electric field. The trends in the exciton g factors entirely reflect the trends of the hole g factors, whereas the electron g factor only shifts the absolute values slightly. To illustrate the resonant behavior of the exciton g factor, the insets in Fig. 5.3(c), mark the localization of the hole states in the coupled dots for the different electric field regimes. As can also be seen from the figure, our calculations are in excellent agreement with the experimental results from Ref. [103].

5.3.2 Giant g factor switching

We now show that the hole g factors in coupled quantum dots can be electrically tuned by as much as 800% and effectively switched between almost zero and a finite value by applying a constant magnetic field in the growth plane rather than along the growth axis. Importantly, we find an unusually pronounced magnetic field anisotropy of the Zeeman splitting within the growth plane. As will be shown below, this is caused by the piezoelectric polarization associated with quantum dot molecules. In Fig. 5.4, the calculated molecular hole g factors for magnetic fields oriented along the planar directions [110] and $[1\bar{1}0]$ are shown. Since the electron g factors remain constant ($g_e = 0.45$) over the shown electric field range, the figure applies to exciton g factors as well. Analogous to the situation for vertically applied magnetic fields, we obtain a resonant reduction and enhancement of the g factors for the bonding and antibonding states, respectively. The molecular ground state [shown in Fig. 5.4(a)] is still localized in the upper and lower dot for large positive and negative electric field, respectively.

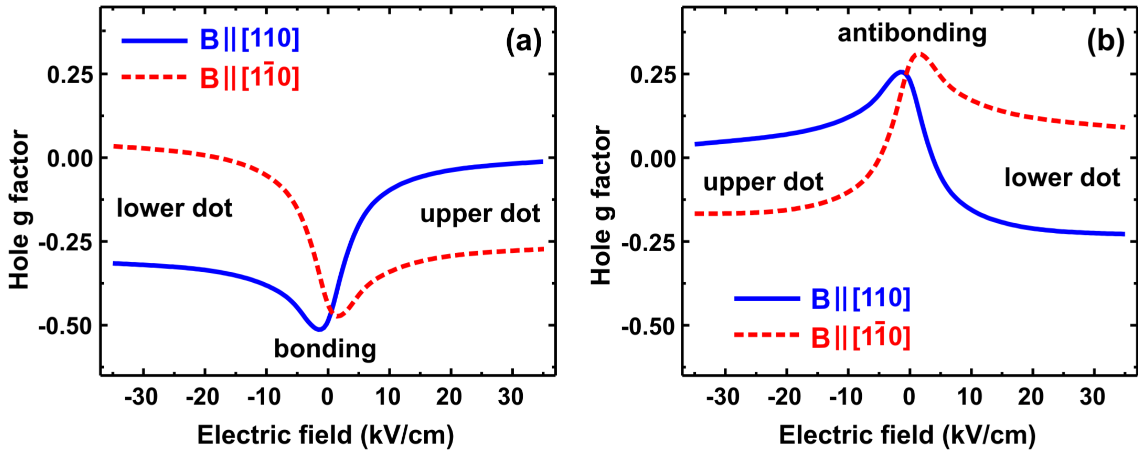


Figure 5.4: (a) Calculated molecular g factors of hole ground state as a function of vertically applied electric field (relative to the resonance field $\mathbf{F}_{\text{res}}$). The constant lateral magnetic field lies along the [110] (full line) and $[1\bar{1}0]$ direction (dashed line), respectively. (b) Same for first excited hole state. For large magnitudes of the electric fields, the molecular hole states are localized predominantly in either the lower or the upper dot, as indicated in the figure.

In fact, we find the character of this state to be almost independent of the modulus and direction of the magnetic field up to 10 T. For the first excited state, the role of the upper and lower dots is reversed.

The results indicate a highly efficient bias induced switching of the molecular g factor between a value close to zero and a finite negative value for the extremal electric field values. This switching behavior is a robust effect and neither sensitive to small changes in the bias, as can be deduced from Fig. 5.4, nor to small changes in the quantum dot widths. Importantly, the calculations predict a pronounced anisotropy of the ground and excited state hole g factors. By orienting the magnetic field along the $[1\bar{1}0]$ instead of the [110] direction, the results indicate that the role of the upper and lower dot in the molecular states are effectively swapped. In summary, we find the following relations to hold for the hole ground state,

$$|g^{[110]}_{\mathbf{E}\to+\infty}| \approx |g^{[1\bar{1}0]}_{\mathbf{E}\to-\infty}| \ll |g^{[1\bar{1}0]}_{\mathbf{E}\to+\infty}| \approx |g^{[110]}_{\mathbf{E}\to-\infty}|. \tag{5.4}$$

Analogous relations can be given for the first excited hole state. Since the g factors associated with the magnetic field directions [110] and $[1\bar{1}0]$ cross each other at the electric field $\mathbf{F}_{\text{res}}$, the Zeeman splitting becomes isotropic within the growth plane for this particular field value. To the best of our knowledge, such a tunable and pronounced anisotropy of quantum dot related g factors has not been observed or discussed so far. In-plane anisotropies of electron g factors in single quantum dots have been found to be quite small as expected [87]. The present effect is caused by

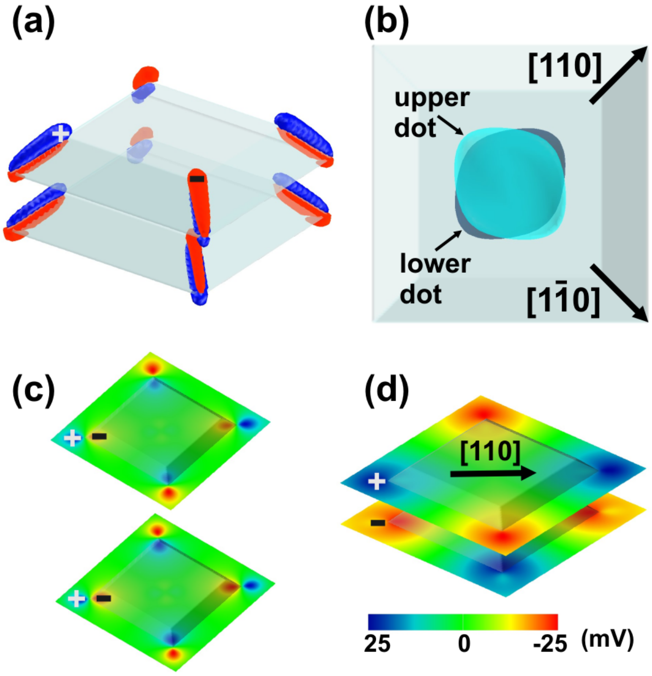


Figure 5.5: (a) Isosurface of calculated piezoelectric polarization charge densities of magnitude $\pm 1.5 \times 10^{19}$ cm^{-3} near the overgrown quantum dots. Positive and negative charges are indicated in blue and red, respectively. (b) Isosurface of probability density of molecular hole ground state for an applied bias of +30 (light blue) and −30 kV/cm (dark blue), respectively. The isosurface is chosen at 25% of the maximum density and the dot separation amounts to $d = 1.5$ nm. For these electric fields, the light blue and the dark blue state is localized in the upper and the lower dot, respectively. (c) Cross sections of the calculated electrostatic potential that results from the piezoelectric charges for a dot separation of $d = 24$ nm. The cross sections are taken at half of the dot height. (d) Same as (c) for $d = 1.5$ nm.

the piezoelectric charge distribution near the edges of the quantum dots that deform the hole charge distribution differently in the two coupled quantum dots that form the molecule. In the following, we will explain this effect in some detail.

For a truncated pyramid shaped buried quantum dot, there are piezoelectric charge dipoles located at the pyramidal edges [109]. The presently calculated piezoelectric charge distributions are shown in Fig. 5.5(a). These dipole charges lead to a potential profile that elongates the hole wave functions along one diagonal and compresses them along the perpendicular direction. For two vertically well separated quantum dots, this potential profile is very similar for both dots. In Fig. 5.5(c), we plot an overlay of two $2D$ cross sections of the electrostatic potential onto the geometric dot profiles. The cross sections are taken at half of the quantum dot height in each dot. The dot separation is $d = 24$ nm in this case. For a smaller dot separation, the multipole potential changes its character completely as shown in Fig. 5.5(d) for $d = 1.5$ nm. The individual quantum dot potential profiles are now rotated with respect to each other by 90°. In addition, the dot molecule behaves as a single vertical dipole at each corner of the pyramid rather than as two consecutive dipoles as in Fig. 5.5(c). In Fig. 5.5(b), we show a top view of the two hole wave functions that lie in different quantum dots. They show a deformation that corresponds to the dipolar potential. The upper and the lower hole wave function is elongated along the $[1\bar{1}0]$ and the [110] direction, respectively. The in-plane g factors in Fig. 5.4 follow this trend since they correlate with the spatial extent of the wave functions. This explains the anisotropy and the switching behavior of the ground and excited state hole g factor.

5.3.3 Efficient universal spin-qubit gate

In this section, we discuss an application of the strong magnetic field anisotropy together with the strong electric field dependence of the g tensor in quantum dot molecules. These two combined effects allow one to fully control the spin precession axis and therefore the spin polarization of a single hole in a quantum dot molecule. There are two prerequisites to achieve such a universal spin-qubit gate. The first one is a static magnetic field along a direction of the dot molecule that is *not* equal to a principal axis of the hole g tensor. The other condition is a selective electric top gate that acts on the quantum dot molecule and applies a bias across the two coupled quantum dots. For individual quantum dots and electrons, such a full Bloch sphere control of the spin polarization has been predicted previously [110]. However, we find the modulation of hole g factors in quantum dot molecules to be an order of magnitude larger than in individual quantum dots. Experimentally, this effect has only been observed in heterostructures for an ensemble of carriers so far [101]. In addition, we find the g factors to form plateaus in their electric field dependence, making the spin manipulation less sensitive to slight variations in the applied voltage pulses.

We can write the Hamiltonian (4.10) in the form

$$H_h = \boldsymbol{\sigma} \cdot \boldsymbol{\Omega}_h, \tag{5.5}$$

where $\boldsymbol{\Omega}_h = 1/2\mu_B \hat{\mathrm{g}}_h \cdot \mathbf{B}$ denotes the spin precession vector and the g tensor of the quantum dot molecule can be given explicitly in the Cartesian basis as

$$\hat{\mathrm{g}}_h = \begin{pmatrix} \frac{\mathrm{g}_h^{[110]} + \mathrm{g}_h^{[1\bar{1}0]}}{2} & \frac{\mathrm{g}_h^{[110]} - \mathrm{g}_h^{[1\bar{1}0]}}{2} & 0 \\ \frac{\mathrm{g}_h^{[110]} - \mathrm{g}_h^{[1\bar{1}0]}}{2} & \frac{\mathrm{g}_h^{[110]} + \mathrm{g}_h^{[1\bar{1}0]}}{2} & 0 \\ 0 & 0 & \mathrm{g}_h^{[001]} \end{pmatrix}, \tag{5.6}$$

where $\mathrm{g}_h^{[110]}$ etc. are the elements of the g tensor along the principal axes. If a magnetic field is applied along the [100] direction, a single spin in this field will start to process

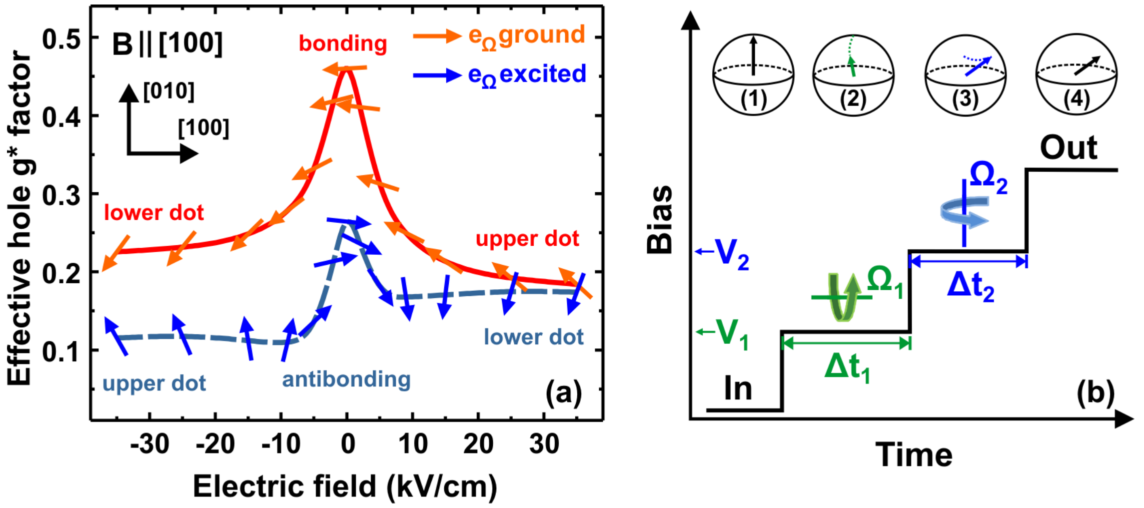


Figure 5.6: (a) Calculated effective g_h^* factors and directions of spin precession axes (arrows) of hole ground state (solid line) and first excited state (dashed line) as a function of the vertical electric field. The constant magnetic field lies in the horizontal [100] direction. (b) Sketch of a puls sequence that allows to coherently manipulate a single spin in a static magnetic field using two bias-controlled orthogonal precession axes $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$, as discussed in the main text. The insets (1)-(4) indicate spin polarizations on the Bloch sphere at different times.

around the axis $\mathbf{e}_\Omega$, which is given by

$$\mathbf{e}_\Omega = \frac{\mathbf{\Omega}_h}{|\mathbf{\Omega}_h|} = \frac{1}{g_h^*}\left(\frac{g_h^{[110]} + g_h^{[1\bar{1}0]}}{2}, \frac{g_h^{[110]} - g_h^{[1\bar{1}0]}}{2}, 0\right). \tag{5.7}$$

The unit vector $\mathbf{e}_\Omega$ is normalized by the effective hole g factor $g_h^* = |\hat{g}_h \cdot \mathbf{B}| / B$. This spin precession axis can be controlled by an applied vertical electric field. In Fig. 5.6(a), we plot the directions $\mathbf{e}_\Omega$ together with g_h^* as a function of the electric field, for the molecular hole ground state and the first excited state. For the molecular ground state, the electric field is able to rotate the precession axis $\mathbf{e}_\Omega$ by 90° from $[\bar{1}\bar{1}0]$ for large negative field magnitudes to the $[\bar{1}10]$ direction for large positive values. This pronounced tunability of $\mathbf{e}_\Omega$ can be deduced from Eq. (5.7) and Fig. (5.4) and is caused by the fact that one of the g factors associated with the magnetic field directions [110] and $[1\bar{1}0]$ vanishes for large magnitudes of the electric field values. For the first excited state, the axis can be rotated even by 270°. In both cases, the modulus of the g factor lies between those of the principal axes in Fig. 5.4 for all electric field values.

In order to fully control the spin of a carrier that is confined to a quantum dot molecule, two orthogonal precession axes are needed. These axes generate the two rotation angles that define any point on the Bloch sphere. For the present structure, one could use, e.g., the axes in Fig. 5.6 that can be associated with large positive and large negative electric field values, respectively. In Fig. 5.6(b), we have sketched this idea using two orthogonal precession vectors $\mathbf{\Omega}_1$and $\mathbf{\Omega}_2$ that are assumed to come with the bias values V_1 and V_2, respectively. We now consider a single hole in a quantum dot molecule with its spin initially polarized along the direction indicated in inset (1) of the figure. By subsequently applying a series of voltage pulses V_1 and V_2, the spin polarization will rotate around the precession vectors $\mathbf{\Omega}_1$and $\mathbf{\Omega}_2$, respectively. The acquired angles indicated in the insets (2) and (3) depend on the time periods Δt_1 and Δt_2. In total, the spin polarization will be coherently manipulated to the final direction in inset (4) that is uniquely determined by the chosen puls sequence. Therefore, this concept allows universal quantum gating with a single electric gate on top of the structure and a static magnetic field.

5.3.4 Resonant electron g factors

We have repeated the calculations in Sec. 5.3.1 for electrons, but assumed a much larger interdot distance $d = 10$ nm in order be able to localize a single electron in each dot separately for large positive and negative electric fields. For the magnetic field lying in the [001] direction, we also obtain a resonant reduction and enhancement of the electron g factors for the bonding and antibonding states, respectively. However, as can be deduced from Fig. 5.7, the effect is about two orders of magnitude smaller than for holes (cf. Fig. 5.3). For the in-plane directions [110] and $\left[1\bar{1}0\right]$, the resonance effect is totally negligible. Even the addition of aluminum to the barrier material (which has been proposed in Ref. [103]) did not increase the effect, in spite of a larger difference between the g factors ($g_e\left(AlAs\right) = 1.52$, while $g_e\left(GaAs\right) = -0.44$). The addition of indium ($g_e\left(InAs\right) = -14.8$) to the barrier only slightly enhances the resonance (about a factor of two for 20% In and $d = 15$ nm), simply because the delocalization of the electron state (in comparison to the hole state) renders its wave function fairly insensitive to reasonable electric fields. Thus, we conclude that electrons are less suitable for this type of g factor engineering in materials with small effective mass. This is in full accord with recent experimental investigations on similar structures performed by Emily Clark from the group of Prof. Jonathan Finley.

5.4 Summary

In summary, we have theoretically investigated electron, hole, and exciton g tensors of vertically stacked quantum dot molecules in vertical and lateral magnetic fields as a function of an applied vertical electric field. We are able to quantitatively explain the experimentally observed [103] resonant enhancements of the g tensor components for vertical magnetic fields without any fitting parameters. For magnetic fields lying in the base plane of the quantum dots, we predict a very pronounced anisotropy in the hole g factors for [110] and $\left[1\bar{1}0\right]$ magnetic field directions. In addition, we predict

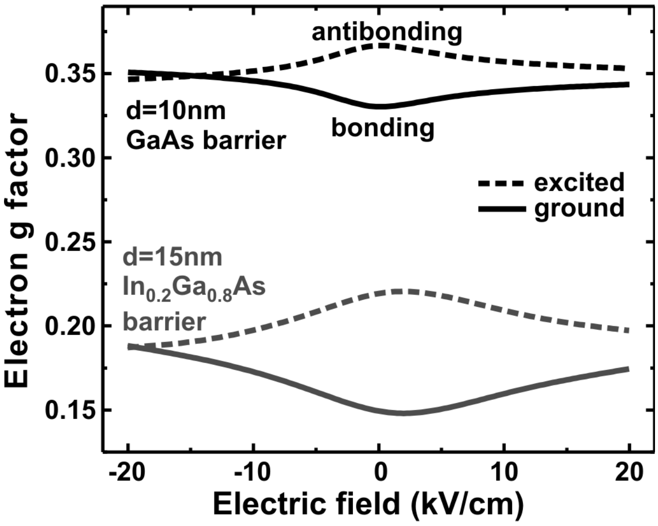


Figure 5.7: Calculated molecular g factors of electron ground state (solid lines) and first excited state (dashed lines), for magnetic and electric fields lying in the vertical [001] direction. The graph shows results for a dot separation of $d = 10$ nm with a pure GaAs barrier (black lines) and a dot separation of $d = 15$ nm with the barrier between the dots consisting of $\mathrm{In}_{0.2}\mathrm{Ga}_{0.8}\mathrm{As}$ (gray lines).

a bias induced g factor switching between almost zero and a finite value for constant magnetic field. Both effects are caused by the piezoelectric charges at the edges of the quantum dots that deform the eigenstates. This tunable anisotropy allows a full control of the spin polarization of a single hole in a quantum dot molecule by a gate voltage and thus the construction of a universal spin-qubit gate.

Chapter 6

Optoelectronic properties of broken-gap heterostructures

6.1 Introduction

The recent fabrication of high quality and high mobility antimonide heterostructures has revived the interest in this material class since their widely tunable electronic structure and optical properties offer many optoelectronic and electronic applications [111–114]. In combination with InAs, antimonides can form type-II heterostructures with a broken gap band alignment. These types of heterostructures are particularly relevant for the development of infrared lasers and infrared detectors [3, 115, 116]. When a material has a broken gap, the strong hybridization of conduction and valence bands [117] prevents an unambiguous separation into electron and hole states [118]. In this situation, the standard multiband EFA fails since this method assumes that electron and hole states can be occupied independently (cf. Sec. 1.3.1). Consequently, only a few theoretical approaches have been developed so far to predict the electronic structure of mesoscopic nanostructures with broken gaps [116, 118–121]. However, a consistent solution for the ambiguity in the calculation of free-carrier charge densities has not been developed so far. Typical models consider only narrow layer structures where one can still rely on the standard prescription of the density calculation [116], or exclude the kinetic coupling between conduction- and valence bands in a single-band approach and therefore miss any nonparabolicities and anisotropies of the band structure [119]. In an earlier introduced approach, only certain components of electronic states are occupied [118]. However, the paper does not give a rigorous justification for this approximation. Recently, another eight-band $\mathbf{k}\cdot\mathbf{p}$ method has been developed that is used explicitly in situations where standard multiband EFA fails [120, 121]. While the details of the density calculation are not specified in the paper, we are absolutely certain that this method differs fundamentally from our approach. This can be deduced from Fig. 1 of Ref. [120], where electron and hole charge densities are given separately, while in our method, as a matter of principle, we can only obtain the sum of both densities.

In the following, we will present a novel electronic structure scheme for broken-gap materials that maintains the efficiency of a continuum approach yet does not depend on an a-priori classification into electron or hole states. In fact, we rely on the envelope function method but turn away from the standard to separate the electrons and holes, and remain in the electron picture throughout. With this procedure, we calculate optical transition energies of intrinsic InAs/GaSb superlattices as a function of the layer thickness. We compare our results with experimental data for narrow layer superlattices where the electron-like and hole-like subbands are ordered as in a conventional semiconductor [115, 122]. In addition, we investigate wide layer superlattices with a strong hybridization between electron-like and hole-like subbands near the Fermi energy. We obtain excellent agreement between theory and experiment [114] and partly revise previous interpretations that lacked a consistent theoretical model.

The chapter is organized as follows. In Sec. 6.2.1, we present our novel charge self-consistent **k·p** envelope function method. The failure of the standard prescription of the charge density calculation is expressed in Sec. 6.2.2 for the situation of broken-gap band alignments. In Sec. 6.2.3, we discuss the individual charge contributions in our novel scheme and consider situations with and without a significant charge transfer between the different layers. We apply the procedure to a series of InAs/GaSb superlattices with different layer widths in Sec. 6.3. Here, we also compare calculated subband energies with experimental data.

6.2 Self-consistent multiband envelope function approach for broken-gap heterostructures

6.2.1 Description of novel method

Broken-gap heterostructures are characterized by the lowest conduction band in one material layer to energetically fall below the highest valence band in an adjacent layer. This leads to a pronounced coupling of bands that needs to be taken into account properly if one wishes to calculate the electronic structure of such systems. Specifically, we consider multi-quantum well systems that are structured along the growth axis z and homogeneous laterally. We rely on the multiband **k·p** envelope function method [11, 13, 16, 19] that we solve in a discrete real-space basis, embracing N grid nodes along the z-axis. In the lateral directions, the wave vector $\mathbf{k}_{\|}$ remains a good quantum number. Our method is valid for any multiband model that includes conduction as well as valence bands. The Schrödinger equation can be written schematically in the form [Eq. (1.41)]

$$\sum_{\mu}\left[\widehat{\mathcal{H}}^{\nu\mu}\left(z,-\mathrm{i}\partial/\partial z,\mathbf{k}_{\|}\right)-e\phi\left(z\right)\right]F_i^{(\mu)}\left(z,\mathbf{k}_{\|}\right)=E_i\left(\mathbf{k}_{\|}\right)F_i^{(\nu)}\left(z,\mathbf{k}_{\|}\right)\ , \tag{6.1}$$

with a Hamiltonian $\widehat{\mathcal{H}}$ that has been patched up from the bulk $\mathbf{k}\cdot\mathbf{p}$ Hamiltonians $\widehat{\mathcal{H}}_{\mathrm{k}\cdot\mathrm{p}}(\mathbf{k})$ of each constituent material, and where the growth direction z has been Fourier transformed into real space ($k_z \rightarrow -\mathrm{i}\partial/\partial z$), as discussed in detail in Sec. 1.2.1. The indices ν, μ run over the N_C conduction bands and the N_V valence bands of a N_A-band $\mathbf{k}\cdot\mathbf{p}$ model with $N_A = N_C + N_V$. $F_i^{(\nu)}$ is the ν-th component of the i-th subband envelope function ($\nu \in \{1, ..., N_A\}$, $i \in \{1, ..., N_A N\}$). The remaining transversal wave vectors are restricted to the two-dimensional Brillouin zone Ω_{BZ} in the reciprocal $\mathbf{k}_\parallel$ space. In order to take into account the spatial charge distribution, the general $N_A \cdot N$ dimensional Hamiltonian is augmented by the electrostatic potential $\phi(z)$ which is determined by the Poisson equation [Eq. (1.65)]

$$\frac{\partial}{\partial z}\varepsilon(z)\frac{\partial}{\partial z}\phi(z) = -\rho(z). \tag{6.2}$$

Here, $\varepsilon(z)$ denotes a position dependent dielectric constant. Since the total charge density ρ depends on the eigenstates resulting from the Schrödinger equation, both Schrödinger- and Poisson equation have to be solved in a self-consistent manner (cf. Sec. 2.6.1). For the concrete applications in this chapter, we use a relativistic eight-band $\mathbf{k}\cdot\mathbf{p}$ envelope function method that has been described in detail in Sec. 1.2.1. Here, strain effects are incorporated into the Hamiltonian via linear band-edge deformation potentials (cf. Sec. 1.4.2). Note that in this model, we have $N_C = 2$ and $N_V = 6$.

In broken-gap heterostructures, the standard envelope function approach (EFA) introduced in Sec. 1.3.1 fails to determine the charge density. This will be shown explicitly in Sec. 6.2.2. We have developed a novel method for the calculation of the free-carrier charge density in multiband $\mathbf{k}\cdot\mathbf{p}$ models that accounts for the problem of overlapping bands and may be called full-band envelope function approach (FB-EFA). In this method, we do not need to classify eigenstates into electron-like or hole-like. Instead of that, we occupy all eigenstates (including the states that have energies in the energy range of the valence bands) with electrons according to the Fermi-Dirac statistics. Afterwards, we subtract an appropriately calculated positive background ionic charge that guarantees charge neutrality. In this approach the free-carrier charge density $\rho = \rho_{\mathrm{FB-EFA}}$ (doping and fixed charges can be added straightforwardly) is given by

$$\rho_{\mathrm{FB-EFA}}(z) = e\left[-n_{\mathrm{fb}}(z) + \rho_{\mathrm{bg}}(z)\right], \tag{6.3}$$

where $n_{\mathrm{fb}}(z)$ is the density of electrons in all included bands (conduction and valence bands) and $\rho_{\mathrm{bg}}(z)$ is the positive background charge density. Concretely, the full-band density n_{fb} is given by,

$$n_{\mathrm{fb}}(z) = \sum_{i=1}^{N_A N} \frac{1}{(2\pi)^2} \int_{\Omega_{BZ}} d^2\mathbf{k}_\parallel \left|F_i\left(z, \mathbf{k}_\parallel\right)\right|^2 f\left(E_i\left(\mathbf{k}_\parallel\right)\right). \tag{6.4}$$

Here, the sum runs over all $N_A \cdot N$ subbands of the Hamiltonian and $f(E)$ is the Fermi function.

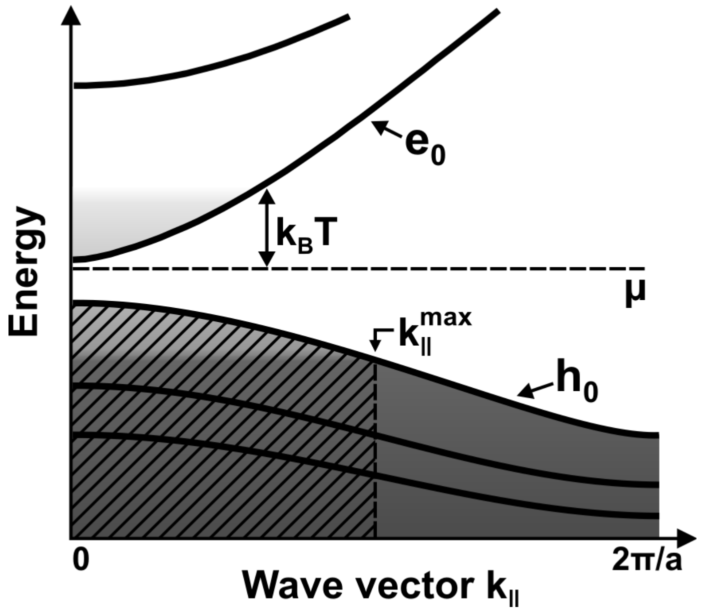


Figure 6.1: Subband dispersion of a narrow gap superlattice, as a function of $\mathbf{k}_{||}$ in an in-plane direction. Regions of fully and partly occupied states are marked in dark and light gray, respectively. The shaded region indicates the required amount of background charge density ρ_{bg}, when the k-space integration is restricted to $k_{||}^{\max}$.

In order to make the total device charge neutral, we subtract the background charge density

$$\rho_{\text{bg}}(z) = N_V N \frac{\Omega_{BZ}}{(2\pi)^2 \Delta z}. \tag{6.5}$$

This density corresponds to the charge of $N_V \cdot N$ fully occupied valence subbands, which is assumed to be homogeneously distributed in order to represent the positively charged ionic cores. The background charge density must be divided by the grid spacing Δz, since the norm of any subband wave function is given by $||F_i|| = 1/\sqrt{\Delta z}$ in a discrete z-space. The two-dimensional Brillouin zone area is given by $\Omega_{BZ} = 8\pi^2/a^2$, where a is the lattice constant. This FB-EFA concept has been adapted from charge self-consistent tight-binding theory [123].

We will now illustrate our novel density model for a superlattice that is assumed to have the conventional ordering of electron- and hole-like subbands but a narrow effective band gap. In Fig. 6.1, we have sketched the energy dispersion of the respective subbands that lie closest to the chemical potential μ, as a function of the $\mathbf{k}_{||}$ wave vector in an in-plane direction. In this structure and for finite temperatures, only the lowest conduction subband (e_0) is partly filled with electrons and the highest valence subband (h_0) is partly filled with holes due to intrinsic Fermi statistics. This has been indicated in the figure by the light gray areas. The filling of states with electrons and holes is performed explicitly in the traditional EFA. In FB-EFA, by contrast, we fill all states with electrons irrespective of their energetic position in the conduction or valence bands. According to Eq. (6.4), the states in the dark gray areas in Fig. 6.1 are fully occupied and those in the light gray areas are partly occupied for finite temperatures due to the characteristics of the Fermi function. Importantly, it is not necessary to integrate the density in Eq. (6.4) over the full two-dimensional Brillouin zone Ω_{BZ}. States that exceed the chemical potential by more than a few k_BT hardly contribute to the sum in Eq. (6.4), since the Fermi function drops exponentially for energies that are larger than μ. Similarly, states that significantly fall below the chemical potential are compensated by the same amount of background charge and do not contribute either. For the antimonide based broken-gap systems considered in this chapter, we find that the wave vectors of partially filled subband states lie within the inner 10% of the Brillouin zone. Thus, we can reduce the maximum wave vector modulus used in Eq. (6.4) to a small value $k_\parallel^{\max}$ so that the number of numerical matrix diagonalizations can be strongly reduced by two orders of magnitude. This requires an additional adjustment of $\Omega_{BZ} \rightarrow \Omega^*_{BZ} = 4(k_\parallel^{\max})^2$ in the relation [Eq. (6.5)] for the background charge density. In FB-EFA, the occupation of the conduction subband is equivalent to the situation in EFA. The light gray valence area, however, is now largely filled with electrons rather than having a minor occupation with holes. In order to obtain the excess charge density ρ, we finally subtract a background charge density ρ_{bg} [Eq. (6.5)] that corresponds to the shaded valence band area in the figure. In Sec. 6.2.3, we will show that both approaches are in fact equivalent for this structure.

We would like to mention an additional computational issue that has been considered in the present implementation. In fact, filling up all states with electrons and subtracting a positive background charge is nominally equivalent to filling all states with holes and subtracting an appropriate negative background charge. This is due to the completeness relation of the eigenstates of a Hermitian Hamiltonian. This relation leads to a spatially homogenous density distribution, when all eigenstate densities are summed up. Since the eight-band model includes less conduction bands ($N_C = 2$) than valence bands ($N_V = 6$), the latter approach is computationally more efficient. Although, we have actually implemented this filling with hole states, we keep with the electron picture in the present discussion to avoid any confusion.

For the concrete predictions presented in this chapter, we consider intrinsic InAs-GaSb (001)-superlattices that have been fabricated and studied experimentally [114]. The structures are characterized by the InAs and GaSb layer widths w_1 and w_2, respectively. Note that we consider only a single period and use periodic boundary conditions in the calculations. The GaSb layers are assumed to be unstrained so that the InAs layers exhibit a slight tensile strain of 0.5%. In Fig. 6.2, we plot the qualitative band edges of the Γ conduction band E_C (dotted line) and the heavy hole valence band E_V (solid line) for two different regimes of layer widths. The graphs show a type-II broken-gap band alignment with an overlap between E_C in the InAs layer

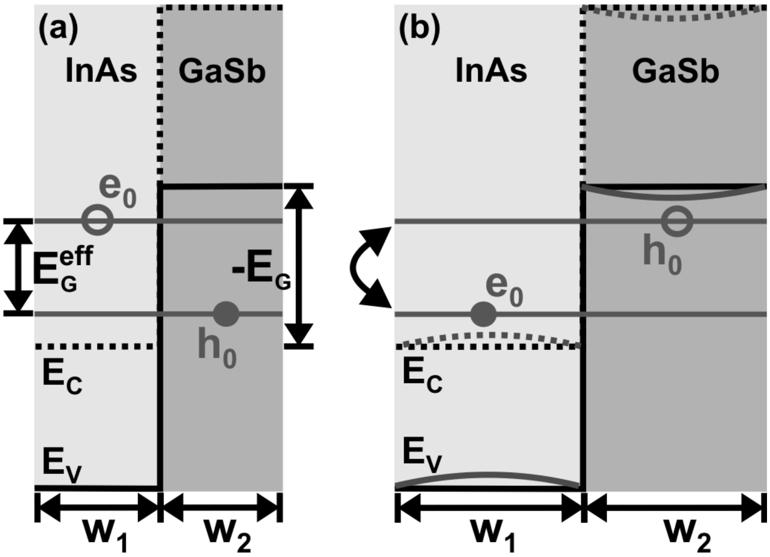


Figure 6.2: Band structure of InAs/GaSb superlattice with InAs and GaSb layer widths w_1 and w_2, respectively. (a) Sketch of the broken gap ($E_G < 0$) band edges in position space with qualitative zone center subband energies in the situation of small layer widths where $E(e_0) > E(h_0)$. (b) Same as in (a) for larger layer widths, where the energetic ordering gets reversed $E(e_0) < E(h_0)$. Here, a charge transfer between the different layers is expected that leads to a band bending as indicated.

and E_V in the GaSb layer ($E_G < 0$). For small layer widths (typically $w_1, w_2 < 10$ nm), confinement can raise the lowest electron subband (e_0) above the highest hole subband (h_0) leading to an ordering of subbands as in a conventional semiconductor [115]. This situation is shown in Fig. 6.2(a). Here, the effective band gap, i.e., the energy difference between the lowest unoccupied and the highest occupied subband at $\mathbf{k}_\parallel = 0$, is given by $E_G^{\text{eff}} = E(e_0) - E(h_0)$. In this situation, electron and hole subbands are well separated in energy so that both the standard EFA, as well as our novel FB-EFA [Eq. (6.3)] can be applied. When moving to larger layer widths, the confinement is reduced, the energetic positions of e_0 and h_0 are switched, and the definition of E_G^{eff} alters to $E_G^{\text{eff}} = E(h_0) - E(e_0)$ [cf. Fig. 6.2(b)]. Now, electron and hole subbands overlap and EFA fails to determine the free-carrier charge densities. Thus, by changing the layer widths, we can effectively tune the present structure between the critical and uncritical situations of standard and reversed ordering of subbands, respectively.

6.2.2 Failure of standard method

In contrast to Eq. (6.3), within the standard EFA, the charge density $\rho = \rho_{\text{EFA}}$ is given by the difference of the electron and hole densities (cf. Sec. 1.3.1)

$$\rho_{\mathrm{EFA}}(z) = e\left[-n_{\mathrm{EFA}}(z) + p_{\mathrm{EFA}}(z)\right], \tag{6.6}$$

with the densities of electrons in the conduction bands

$$n_{\mathrm{EFA}}(z) = \sum_{i=N_V N+1}^{N_A N} \frac{1}{(2\pi)^2} \int_{\Omega_{BZ}} d^2\mathbf{k}_{\|} \left|F_i\left(z, \mathbf{k}_{\|}\right)\right|^2 f\left(E_i\left(\mathbf{k}_{\|}\right)\right), \tag{6.7}$$

and holes in the valence bands

$$p_{\mathrm{EFA}}(z) = \sum_{i=1}^{N_V N} \frac{1}{(2\pi)^2} \int_{\Omega_{BZ}} d^2\mathbf{k}_{\|} \left|F_i\left(z, \mathbf{k}_{\|}\right)\right|^2 \left[1 - f\left(E_i\left(\mathbf{k}_{\|}\right)\right)\right]. \tag{6.8}$$

In Eq. (6.7), our prescription of electron densities [Eq. (6.4)] has been modified in the sum over i that now runs only over the $N_C \cdot N$ energetically highest subbands of the Hamiltonian $[N_A \cdot N = (N_C + N_V) \cdot N]$. Note that we have assumed the subbands to be ordered ascending in energy, so the first $N_V \cdot N$ subbands lie in the valence bands and the remaining $N_C \cdot N$ subbands lie in the conduction bands. However, for multi-band $\mathbf{k}{\cdot}\mathbf{p}$ models that couple conduction- and valence bands (such as the eight-band model), it is in general not possible to decide if a specific eigenstate has to be occupied by an electron or by a hole. This somewhat unexpected situation occurs in nanostructures that do not have a global energy gap as will be discussed in the following.

In order to express the failure of the EFA, we now turn to the case of a reversed ordering of subbands $E(e_0) < E(h_0)$ in Fig. 6.2(b). To this end, we consider a superlattice with an InAs layer width of $w_1 = 16$ nm and a GaSb layer width of $w_2 = 8$ nm. In Fig. 6.3, we show the subband dispersion of this structure, as a function of the wave vector in the [010] direction. In the following, we will only consider the third and fourth subband (in energetically ascending order) from Fig. 6.3 that we will call lower and upper subband, respectively. At some $\mathbf{k}_{\|}$ points, one may assign certain states to have the character of an electron (a hole) if they are almost completely localized in the InAs (GaSb) layer, where the band gap lies below (above) the energies of the states. Namely, at $\mathbf{k}_{\|} = 0$, the lower subband is electron-like since it is localized in the InAs layer (see lower left inset in Fig. 6.3). By contrast, the upper subband is localized in the GaSb layer at $\mathbf{k}_{\|} = 0$ and therefore hole-like. At $\left|\mathbf{k}_{\|}\right| = 0.04 \cdot 2\pi/a$, however, the characters of these two subbands are switched. Thus, it is not possible to assign the character electron- or hole-like to a complete subband, unambiguously. For intermediate wave vectors $\left|\mathbf{k}_{\|}\right| = 0.02 \cdot 2\pi/a$ the band structure shows an anticrossing between the two subbands that leads to a strong mixing of electron and hole states. Indeed, the probability distributions of the eigenstates of the two subbands shown in the insets for a $\mathbf{k}_{\|}$ point near the anticrossing exhibit two distinct maxima that correspond to an almost equal localization in the InAs and GaSb layers. Obviously, these states cannot be classified as electron- or hole-like and effectively contribute to the electron as well as to the hole charge densities. Thus, it is

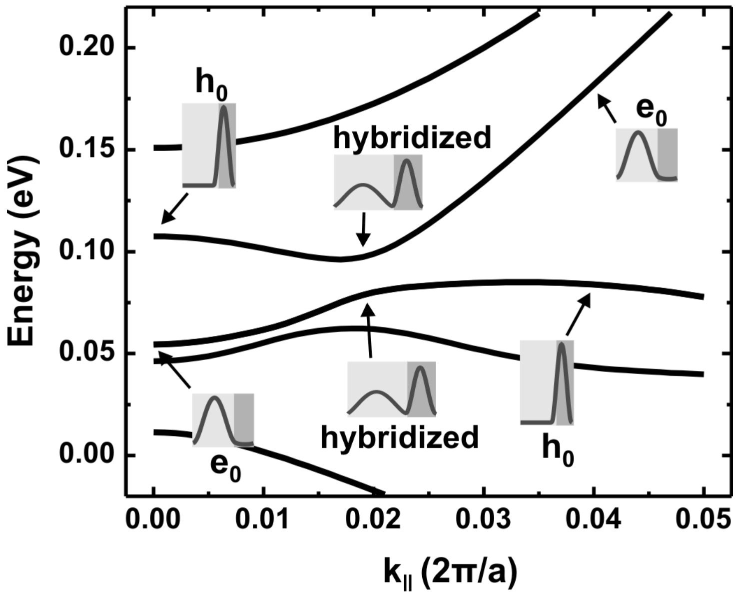


Figure 6.3: Subband dispersions as a function of $\mathbf{k}_{\|}$ in the lateral direction, of an InAs/GaSb superlattice with InAs and GaSb layer widths of $w_1 = 16$ nm, and $w_2 = 8$ nm, respectively. The insets show probability densities of subband eigenstates at three different values of $\mathbf{k}_{\|}$ (0, 0.2, and 0.04 in units of $2\pi/a$, where a denotes the lattice constant of GaSb). The character of the individual eigenstates can be assigned from their localization in the InAs (light gray) or the GaSb (dark gray) layers. All energies are given relative to the zone center conduction band maximum in the InAs layer.

not possible to simply occupy some subbands according to the Fermi-Dirac statistics of electrons and others with the statistics of holes as it is done in the standard EFA of Eq. (6.6).

6.2.3 Discussion of charge density contributions

Next, we will discuss the individual charge density contributions that are obtained within our novel FB-EFA. In order to critically assess this method, we start with the situation of narrow layer widths from Fig. 6.2(a). Concretely, we consider an intrinsic superlattice characterized by the layer widths $w_1 = w_2 = 9$ nm at $T = 0$ K where EFA yields zero density throughout the structure. In Fig. 6.4(a), we plot the total charge density for the InAs/GaSb superlattice that has been determined within FB-EFA. Here, the coupled eight-band Schrödinger equation [Eq. (6.1)] and the Poisson equation [Eq. (6.2)] have been solved iteratively until self-consistency is reached.

The resulting charge density basically consists of dipoles at each of the material layer interfaces. These dipoles are produced not only by the states close to the

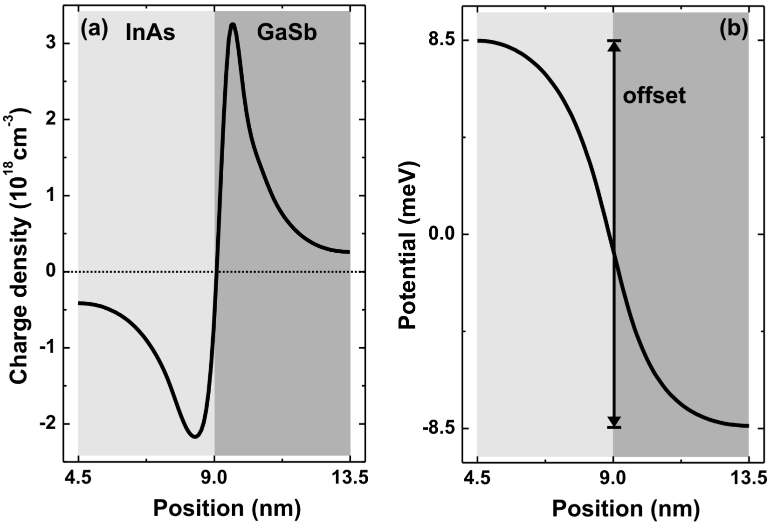


Figure 6.4: (a) Local charge density close to a material interface obtained within FB-EFA for an intrinsic InAs/GaSb superlattice with layer widths of $w_1 = w_2 = 9$ nm. (b) Appropriate potential profile $V(z) = -e\phi(z)$ that results from solving the Poisson equation [Eq. (6.2)] with the charge density from (a). As can be seen, the interface dipoles effectively introduce a jump in the potential that modifies the offset between the valence band edges.

chemical potential, but also by lower lying states, whose density is redistributed with respect to the bulk by the insertion of a material interface. Such interface dipoles are well known from first-principles electronic structure calculations and are in fact the major origin for the band-edge discontinuities between different materials [124]. In the present method, the dipole charges are quantitatively incorrect, but this problem is a removable one. The incorrectness comes from the fact that large-k states strongly contribute to the dipole charges. Such high energy states are beyond the validity of $\mathbf{k}\cdot\mathbf{p}$ theory. The envelope function method yields accurate eigenstates only for small values of $\mathbf{k}$, because it is based on an expansion of the band structure around the zone center. However, in the solution of the Poisson equation, the interface dipoles result only in a small jump in the electrostatic potential as is shown in Fig. 6.4(b). In the present method, this additional offset is meant to be already included in the total offset between different semiconductor band edges. These band offsets can be determined either from the model solid theory by Van de Walle [125], or in terms of electronic states of the atomic cores as it has been proposed by Wei and Zunger [126]. In our method, we rely on the energy values from the (more recent) latter work. Since the interface dipole induced band offset should already be included in the given offset, it has to be subtracted to avoid double counting. We find that for given k-space restriction $k_{\|}^{\max}$, the size of the interface dipole only depends on the two comprised materials. For the present InAs/GaSb interface with $k_{\|}^{\max} = 0.05 \cdot 2\pi/a$, we obtain a dipole-induced offset of -17 meV, when moving from the InAs layer to the GaSb layer. In all following calculations, we therefore increase the band offset between InAs and GaSb from Ref. [126] by 17 meV. We find that the modification of band offsets itself does not alter the interface charge dipoles significantly and therefore the influence of these dipoles can be removed, effectively.

We would like to mention that one could criticize the present method for filling (lower lying) large-k states that are beyond the validity of the underlying theoretical concept. However, these states are basically needed only to fulfill a completeness relation. For a homogenous material, the sum over all included valence states results in a homogenous charge density that corresponds to ρ_{bg} in Eq. (6.3), irrespective on how accurate the lowest states are. We fill the incorrect states just to compensate ρ_{bg}. In a heterostructure, they are slightly modified close to the interfaces and lead to a charge redistribution. So the incorrect states effectively contribute only to the interface dipoles, whose effect is removed anyway.

Now, we turn to the situation of wide layer widths $w_1 = 18$ nm and $w_2 = 9$ nm. Fig. 6.5(a) shows the subband dispersion for the same situation as in Fig. 6.3, now calculated in terms of the FB-EFA method. In addition, in Fig. 6.5(b), we show the calculated charge density distribution that can be explained qualitatively by considering the character of the subbands. As has been stated before, the eigenstate at $\mathbf{k}_{\|} = 0$ of the lowest subband above the chemical potential μ originates in the valence bands. Since its energy is larger than μ, its occupation lacks electrons which in turn creates a positive charge contribution. Since the probability density of this state is dominantly localized within the GaSb layer (dark gray area), we obtain a positive charge distribution in this layer. Qualitative similar statements hold for all other wave vectors with $|\mathbf{k}_{\|}| < 0.02 \cdot 2\pi/a$ in this subband, as has been indicated by the open circles in Fig. 6.5(a). Analogously, the zone center eigenstates of the highest subband below μ (filled circles) originate from the conduction band which leads to a negative charge distribution within the InAs layer (light gray area). Altogether, this results in a charge transfer of the order of 10^{18} cm^{-3} between the GaSb and the InAs layers. Besides the interface dipoles, the charge density distribution in Fig. 6.5(b) clearly shows a nonzero base density at the center of each layer which comes from this charge transfer. Note that it would be difficult to accurately calculate such a small charge transfer in an atomistic approach that includes all electronic states in the Brillouin zone and therefore deals with charges of the order of 10^{23} cm^{-3}.

6.3 Optical transitions in InAs/GaSb superlattices

In this section, we apply our novel method to calculate optical transition energies of type-II superlattices with a broken gap. Here, we delve further into the properties of the InAs/GaSb superlattices that we already used in the previous sections to exemplify the arising difficulties and to discuss the individual charge contributions.

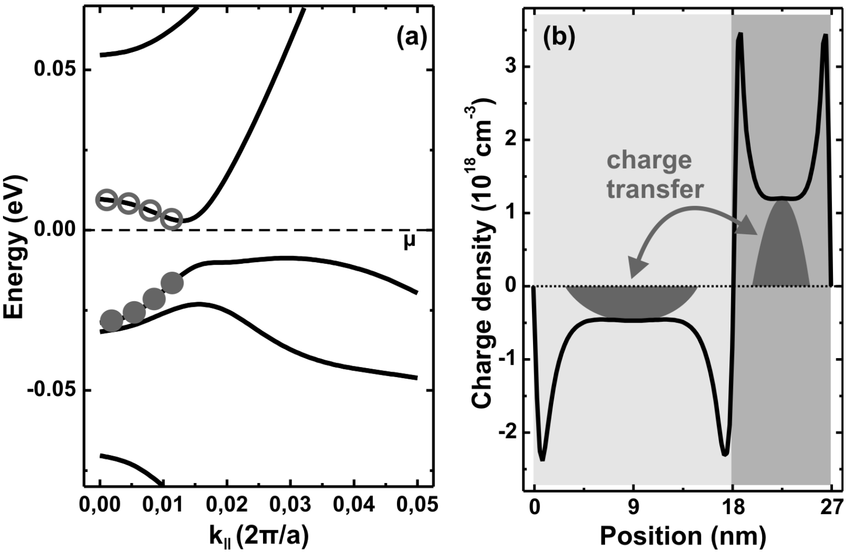


Figure 6.5: (a) Charge self-consistent FB-EFA subband dispersion of intrinsic InAs/GaSb superlattice with InAs and GaSb layer widths of $w_1 = 18$ nm, and $w_2 = 9$ nm, respectively. Occupied electron states and unoccupied hole states are indicated by filled and open circles, respectively. (b) Calculated local charge density obtained within FB-EFA. Besides the interface dipoles, a charge transfer between the InAs and the GaSb layers takes place, as indicated.

6.3.1 Narrow layer structures

For optical applications, very important properties of such superlattices are transition energies, which depend on the relative energies of electron and hole subbands. One of these transition energies is given by the effective band gap E_G^{eff}, which has been defined in Sec. 6.2.1. We have calculated E_G^{eff} for superlattices with InAs and GaSb layer widths that vary absolutely but are hold equal relative to each other ($w_1 = w_2 = w$). Concretely, we have investigated layer widths between 2 and 12 nm, and the temperature has been set to 4 K. As can be deduced from Fig. 6.6, our results show reasonable agreement with experimental values from Refs. [115, 122]. The deviations can be explained by the fact that the present continuum model becomes less accurate for the smallest layer widths considered.

We now turn to a brief discussion of the observed trends in the effective band gap. For the present structure, we are in the regime of Fig. 6.2(a), where strong confinement raises the lowest electron-like subband (e_0) above the highest hole-like subband (h_0) leading to $E_G^{\text{eff}} = E(e_0) - E(h_0)$. With increasing layer widths, the electron-like subbands decrease in energy and the hole-like subbands increase in energy due to the reduction in carrier confinement. Consequently, E_G^{eff} decreases as a function of the layer widths. At $w \approx 10$ nm, E_G^{eff} becomes zero. By further increasing w, the subbands e_0 and h_0 swap their positions so that the definition of the effective

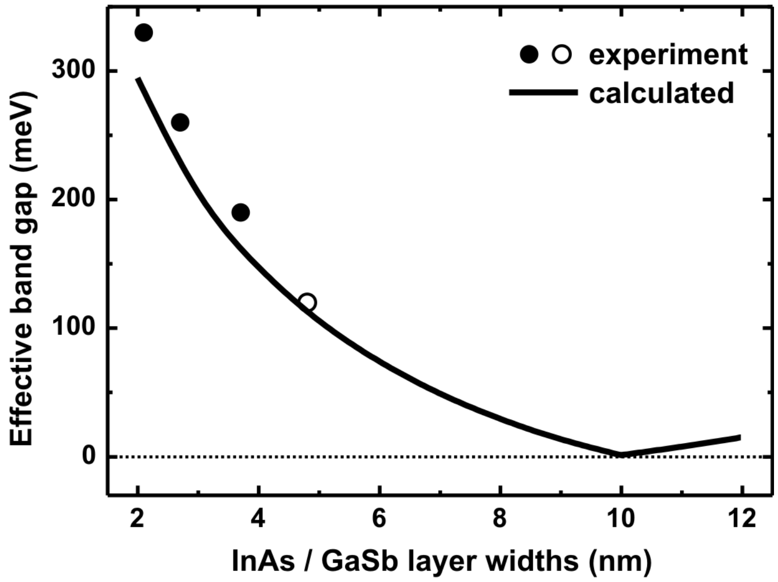


Figure 6.6: Calculated effective band gaps E_G^{eff} of InAs/GaSb superlattices as a function of the layer width $w_1 = w_2$. For comparison, we also show experimental data from Ref. [122] (full circles) and from Ref. [115] (open circle).

band gap alters to $E_G^{\text{eff}} = E(h_0) - E(e_0)$. Thus, E_G^{eff} now increases due to the reduction of confinement. We would like to point out that for $w > 10$ nm, a significant charge transfer between the InAs and GaSb layers occurs (cf. Fig. 6.5) that leads to a band bending [indicated in Fig. 6.2(b)] which shifts the lowest electron and hole subbands upwards and downwards, respectively. In addition, a charge self-consistent calculation is required to determine the chemical potential and therefore the highest occupied and lowest unoccupied subbands.

6.3.2 Wide layer structures

In this section, we investigate the regime of wide layer superlattices $w_1, w_2 \geq 10$ nm in more detail. Here, the energetic order of electron-like and hole-like subbands is altered, the subband dispersion shows anti-crossings and electron and hole states hybridize, making it impossible to distinguish them in general (cf. Fig. 6.3). Using the FB-EFA, we have calculated transition energies between subbands at $\mathbf{k}_\| = 0$ for superlattices with InAs layer widths w_1 in the range of 10 to 55 nm and a fixed GaSb layer width of $w_2 = 10$ nm at $T = 4$ K. Our results are shown in Fig. 6.7. We have assigned the character of each subband by the shape and location of its wave function at $\mathbf{k}_\| = 0$. We find the lowest electron-like subband e_0 to fall below the chemical potential and the higher lying subbands h_0, e_1, e_2 to be unoccupied at $\mathbf{k}_\| = 0$ for the entire range of layer widths considered. With increasing w_1, electron intersubband

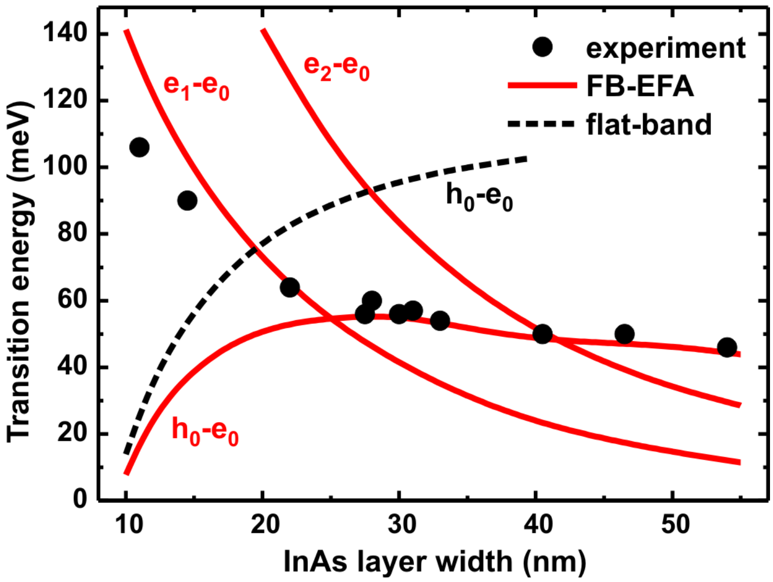


Figure 6.7: Calculated interband and intersubband transition energies of InAs/GaSb superlattices with fixed GaAs layer width $w_2 = 10$ nm, as a function of the InAs layer width w_1. We compare charge self-consistent FB-EFA results (solid line) with a flat-band approximation that does not include free carrier charges (dashed line) and experimental results from Ref. [114] (circles).

transition energies ($E\left(e_1\right) - E\left(e_0\right)$, $E\left(e_2\right) - E\left(e_0\right)$) always decrease due to reduction of confinement. This simple trend does not hold for the interband transition energy $E\left(h_0\right) - E\left(e_0\right)$. For the smallest values of w_1, the interband transition between the electron ground state and the hole ground state increases rapidly in energy with increasing layer width. This is again founded in the reduction of confinement which now increases the energy difference [cf. Fig. 6.2(b)]. For larger layer widths, this transition energy saturates as the charge transfer induced band bending compensates the reduction of confinement. Only for very large layer widths, the charge transfer effect slightly dominates and leads to a reduction of the transition energy.

Our results are in excellent agreement with experimental data [114]. We can confirm that for layer widths $w_1 < 25$ nm, the observed transition is an intersubband transition between electron ground state and first excited state. For larger widths $w_1 > 25$ nm, our results suggest that the observed transition is in fact an interband transition between electron ground state and the higher lying hole ground state. This contradicts the assumption of Poulter et al. [114] to measure only intersubband transitions. However, the slopes of the intersubband transition energies do not fit to the experimental values at all. The measured energies show a pronounced saturation for the largest layer widths that can be reproduced only by the calculated interband transition. Note that the results obtained in a flat-band approximation (which leaves out any potential corrections induced by charges) strongly deviate from the charge self-consistent solution. This clearly demonstrates that a consistent theoretical model for wide layer broken-gap heterostructures is inevitable to understand the experimental results.

6.4 Summary

In Summary, we have developed a novel charge self-consistent multiband $\mathbf{k}\cdot\mathbf{p}$ envelope function method for the calculation of the electronic structure of type-II broken-gap superlattices. When a material has a broken gap, standard multiband $\mathbf{k}\cdot\mathbf{p}$ approaches fail since they depend on an unambiguous separation into electron and hole states. The key point in the presently developed full-band envelope function approach (FB-EFA), however, is to remain in the electron framework throughout. In fact, we occupy all included subbands with electrons according to the Fermi statistics and subsequently subtract a positive background ionic charge that guarantees charge neutrality. With this procedure, we have calculated local charge densities and subband dispersions of InAs/GaSb superlattices. We show that the FB-EFA method is equivalent to the standard EFA when applied to narrow layer structures with an ordering of subbands as in a conventional semiconductor. For wide layer superlattices, the FB-EFA correctly yields the charge transfer between the InAs and GaSb layers. In addition, we have calculated optical transition energies of intrinsic InAs/GaSb superlattices as a function of the layer thickness. We find excellent agreement with experimental data [114] in a regime where the superlattices exhibit a crossover in the energetic order of the lowest electron-like and the highest hole-like subbands.

Chapter 7

Shallow impurity states in silicon

7.1 Introduction

Besides quantum dots that have been extensively discussed in the chapters 4 and 5, there is another interesting system in semiconductors where carriers are subject to three-dimensional confinement, namely impurities. In this chapter, quantitative electronic structure properties of shallow impurities are presented that have been investigated in close collaboration with André Stegner from the experimental group of Prof. Martin Brandt.

Electron paramagnetic resonance (EPR) is known to be an important tool to identify paramagnetic defects in semiconductors and to manipulate the respective impurity spin states. The latter aspect is particularly relevant for silicon based quantum computing. In EPR, a magnetic field is applied that splits the Kramers degenerate impurity spin states. Then, the individual spin transitions can be excited by bringing the system into resonance with a strong microwave field [127]. It is known that the EPR lines of boron in silicon exhibit a pronounced inhomogenous broadening that depends on the orientation of the magnetic field [128]. However, the origin of the dominant Gaussian contribution of the broadening has not been understood so far. Recently, the EPR lineshape of isotopically pure ^{28}Si has been investigated [129], showing a strong reduction of the linewidth. This suggests an isotopic effect to be responsible for the broadening in the natural silicon samples. Similarly, the residual acceptor ground-state splitting in silicon has been successfully explained by means of isotopically induced local fluctuations of the valence band edge [130]. In Sec. 7.2, this approach is extended to nonzero magnetic fields, in order to investigate the isotopic shift of boron acceptor Zeeman levels and its influence on the EPR line-broadening. Further details on this analysis can be found in Ref. [129] together with novel experimental results.

Recently, there has also been a growing interest in freestanding silicon nanocrystals (Si-NCs), due to the novel capability of producing macroscopic amounts of these structures [131]. When thinking of devices based on Si-NCs, the electronic properties can be tailored via doping, which differs significantly from the situation in bulk [132]. The underlying mechanisms such as the dopant confinement and its impact on the ionization energy and doping efficiency are not yet fully understood. So far, available experimental data have been explained on the basis of quantum confinement only [133]. However, it is known that modifications of the electronic states may also result from a suppression of dielectric screening of the Coulomb interaction [134]. In Sec. 7.3, we investigate the influence of quantum confinement as well as dielectric effects on the localization of donor wave functions in Si-NCs of different size. The author's present calculations have been published in Ref. [135], where they are used to interpret recent experimental data.

7.2 Isotopic shift of acceptor Zeeman levels

Boron acceptors in bulk silicon (Si:B) are conventionally described in the hydrogenic impurity model, where the free charge carrier moves in the electrostatic potential of the singly charged impurity ion. The ground-state envelope function of the acceptor hole is the 1s state of a hydrogen-like atom, as shown schematically in Fig. 7.1. In fact, this ground state is fourfold degenerate due to spin and Bloch basis angular momentum. In a magnetic field, the degeneracy of spin-resolved heavy and light holes is removed so that the eigenstates can be approximately classified in the total angular momentum basis $|j = 3/2, m_j \in \{\pm 1/2, \pm 3/2\}\rangle$ (cf. Sec. 1.2.1). Natural silicon ($^{\mathrm{nat}}$Si) is known to consist of the isotopes ^{28}Si (92.23%), ^{29}Si (4.67%), and ^{30}Si (3.1%) that are randomly distributed in a bulk crystal. The influence of different isotopes on the electronic structure can be described in terms of local fluctuations of the band-edge energies [130] (see inset in Fig. 7.1). This leads to shifts of the individual acceptor ground state Zeeman levels which will be investigated in the following.

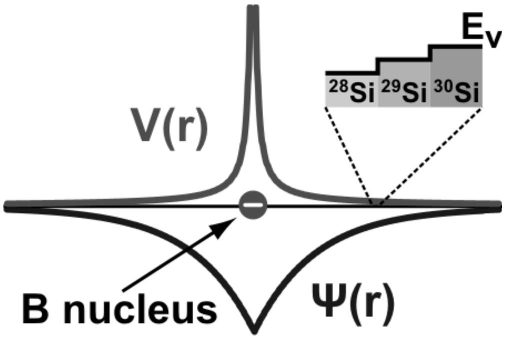


Figure 7.1: Schematic cross section of the boron acceptor ground state in silicon. The hydrogen-like potential $V(r)$ of the acceptor impurity leads to a strong localization of the hole wave function $\Psi(r)$ close to the nucleus. The inset indicates the different valence band energies of the silicon isotopes.

Method

In order to determine isotopically induced modifications of the magnetic energy spectrum, we first calculate the four Zeeman levels of the boron acceptor ground state in a magnetic field and subsequently treat the isotope effect on the valence band edge as a perturbation. We use the six-band **k·p** envelope function model introduced in Sec. 1.2.1 that has been augmented by an external magnetic field as discussed in chapter 3. Furthermore, we need to add a suitable Coulomb potential. The Hamiltonian can be written schematically in the form

$$\hat{H} = \hat{H}^{6\times6}_{\mathbf{k}\cdot\mathbf{p}}\left(\mathbf{r},\mathbf{r}',\mathbf{B}\right) + \frac{\mathrm{g}_0\mu_B}{2}\hat{\mathbf{S}}^{6\times6}\cdot\mathbf{B} + V\left(r\right), \tag{7.1}$$

where the first term on the right-hand side represents the six-band effective mass Hamiltonian [Eq. (1.17)] in a discrete real-space basis. This term includes the coupling to the magnetic field **B** in a nonperturbative and manifestly gauge-invariant manner, with **B** only appearing in phase factors. The second term on the right-hand side of Eq. (7.1) couples the spin to the field **B**. Here, μ_B is the Bohr magneton, $\mathrm{g}_0 = 2$ is the free-electron g factor, and the 6×6 spin matrices $\hat{S}_i = 1^{3\times3}\otimes\hat{\sigma}_i$ $(i\in\{x,y,z\})$ are completely determined by the Pauli matrices $\hat{\sigma}_i$. The impurity nucleus is represented by a negative charge at the center of the simulation domain screened by the bulk silicon dielectric constant $\varepsilon_{\mathrm{Si}} = 11.7$. So, the potential energy of the acceptor hole is given by

$$V\left(r\right) = \frac{e^2}{\varepsilon_{\mathrm{Si}}r} + W\left(r\right). \tag{7.2}$$

We have also included the so-called central cell correction $W(r)$, which describes the potential very close to the nucleus of dopant atoms [136] and is known to strongly affect the impurity energy levels. Here, we rely on the parametrization from Ref. [137]

$$W\left(r\right) = \frac{e^2}{r}\left[A\exp\left(-\alpha r\right) + \left(1-A\right)\exp\left(-\beta r\right) - \frac{\exp\left(-\gamma r\right)}{\varepsilon_{\mathrm{Si}}}\right], \tag{7.3}$$

where the parameters $\alpha = 0.755/a_B$, $\beta = 0.35/a_B$, $\gamma = 2.45/a_B$, and $A = 1.14$ are some fitting parameters. In order to avoid the singularity at $r = 0$, the potentials are replaced by $Q\delta_{r,0}$ for the central grid node, following Ref. [137]. The concrete choice of the parameter Q will be discussed later. In order to account for the long range character of the Coulomb potential together with the strong confinement at the acceptor nucleus, we use a simulation domain size of 50 nm along each axis and an inhomogenous grid with a strong concentration of nodes close to the acceptor position. With this model, we calculate the four ground state Zeeman levels (labeled by $m_j = +3/2, +1/2, -1/2, -3/2$) in external magnetic fields with different orientations relative to the crystallographic axes.

Next, the isotopic modifications of the band edges are considered perturbatively. To this end, we map the wave functions on an atomistic silicon crystal lattice. Here, the envelope function values $\Psi\left(\mathbf{n}\right)$ at the positions of the silicon nuclei are evaluated

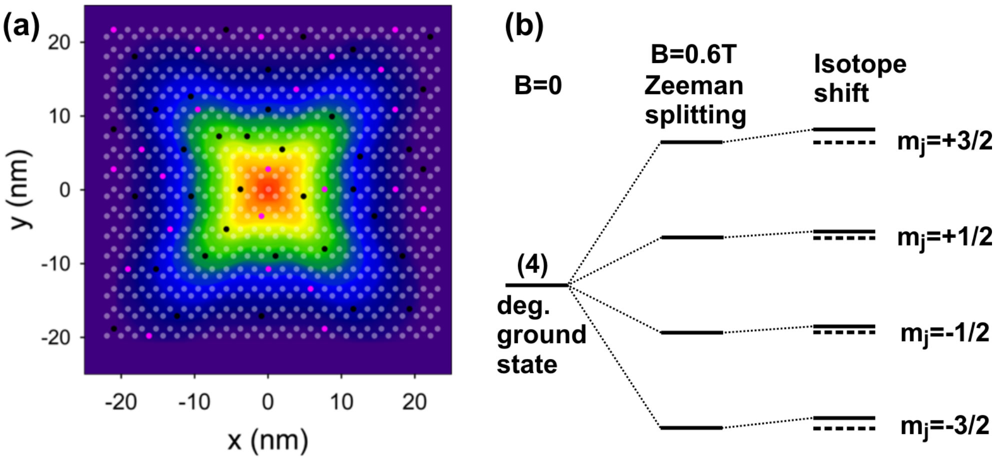


Figure 7.2: (a) Two-dimensional cross section of calculated $m_j = 3/2$ acceptor ground state probability density for a magnetic field of $B = 0.6$ T. The filled circles show the positions of ^{28}Si (light gray), ^{29}Si (black), and ^{30}Si (magenta) nuclei. (b) Schematic modifications of the energy spectrum by subsequent inclusion of the magnetic field and the isotopic perturbation potential.

by linear interpolation. As is shown in Fig. 7.2(a), we choose a random configuration of ^{28}Si, ^{29}Si, and ^{30}Si isotopes under the constraint of their natural accumulations in silicon. Following Ref. [130], we introduce the isotopic perturbation potential

$$V_{\mathrm{iso}}(\mathbf{n}) = \begin{cases} 0 & \text{for } ^{28}\mathrm{Si} \\ \Delta E^{29} & \text{for } ^{29}\mathrm{Si} \\ \Delta E^{30} & \text{for } ^{30}\mathrm{Si} \end{cases} . \tag{7.4}$$

Here, the valence band shifts $\Delta E^{29} = 0.74$ meV and $\Delta E^{30} = 1.46$ meV arise from a renormalization of the energies, introduced by the electron-phonon interaction via the mass dependence of the amplitudes of zero-point fluctuations [130]. Next, we project the diagonal perturbation potential into the subspace spanned by the four Zeeman states $|\Psi_i\rangle$, leading to a 4×4 perturbation Hamiltonian $\hat{H}^{4\times4}_{\mathrm{iso}}$ with the elements given by

$$H^{ij}_{\mathrm{iso}} = \langle \Psi_i(\mathbf{n}) | V_{\mathrm{iso}}(\mathbf{n}) \, \delta_{\mathbf{nm}} | \Psi_j(\mathbf{m}) \rangle . \tag{7.5}$$

Finally, we diagonalize the total Hamiltonian $\hat{H} = \hat{H}^{4\times4}_{\mathbf{J}\cdot\mathbf{B}} + \hat{H}^{4\times4}_{\mathrm{iso}}$ including the Zeeman term $\hat{H}^{ij}_{\mathbf{J}\cdot\mathbf{B}} = E_i \delta_{ij}$, with E_i being the energy of the i-th Zeeman level. The isotopic contribution $\hat{H}^{4\times4}_{\mathrm{iso}}$ mixes the unperturbed eigenstates $|\Psi_i\rangle$ so that the eigenvalues of $\hat{H}$ are slightly shifted with respect to the unperturbed Zeeman levels, as drawn schematically in Fig. 7.2(b). From these eigenvalues, one can extract the respective isotope shifts.

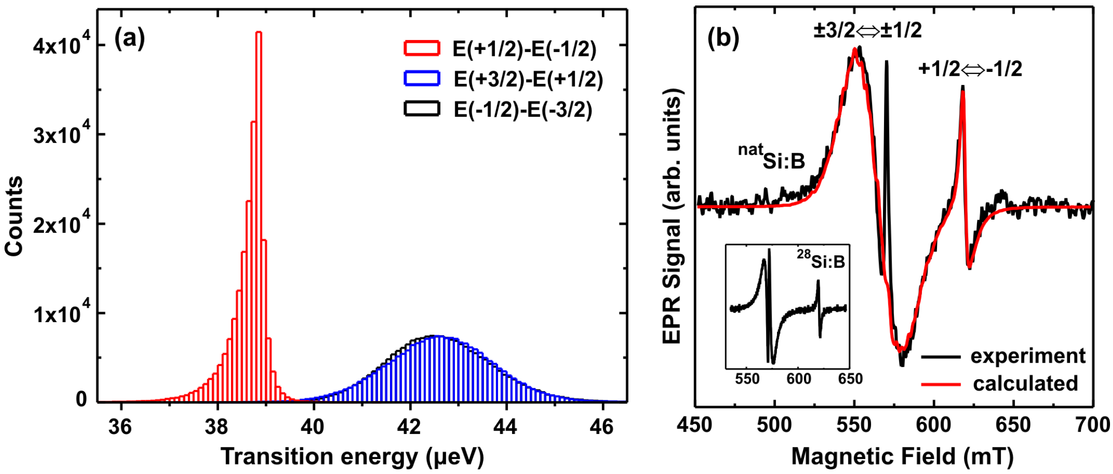


Figure 7.3: (a) Statistical distribution of the $\Delta m_j = 1$ transition energies for a total number of $200,000$ different isotope configurations. The magnetic field of 0.6 T is lying in the [001] direction. (b) Calculated convolution (red line) of distributions from (a) with purely Lorentzian lines that have the linewidths of the experimental EPR signal of isotopically pure ^{28}Si:B (inset). A comparison with the EPR signal of $^{\mathrm{nat}}$Si:B (black line) shows excellent agreement.

Results

We have performed the calculation of the isotope shifts for $200,000$ different random isotope configurations and a magnetic field of 0.6 T. In Fig. 7.3(a), the resulting statistical distribution of the $\Delta m_j = 1$ transition energies is shown for the magnetic field lying in the [001] direction. We find the distribution of the $E\,(+1/2) - E\,(-1/2)$ transition energies to be very narrow, while the $E\,(+3/2) - E\,(+1/2)$ and $E\,(-1/2) - E\,(-3/2)$ transitions vary by up to 10% due to the isotope shifts. This can be understood by the fact that the former transition takes place between two very similar states that are dominantly light-holes and experience very similar isotope shifts. By contrast, the latter transitions comprise light hole like and heavy hole like states which are more differently perturbed by the isotopes.

The statistical distribution of the transition energies can be used to explain the experimentally observed broadening of the EPR lines. In Fig. 7.3(b), we show EPR signals for $^{\mathrm{nat}}$Si:B (black line) and isotopically pure ^{28}Si:B (inset), recently measured by the group of Prof. Kohei Itoh (Keio University, Japan) [129]. The microwave source had a resonance energy of 40 μeV (9.7 GHz). As one can see, the $^{\mathrm{nat}}$Si:B signal exhibits a pronounced broadening with respect to the ^{28}Si:B signal. In order to analyze the isotopic effect, we have convoluted the transition energy distribution functions from Fig. 7.3(a) with purely Lorentzian lines that have the linewidths of the ^{28}Si:B spectra. The resulting calculated signal (red line) shows striking agreement with the $^{\mathrm{nat}}$Si:B signal (when neglecting the fine structure of the left transition), and

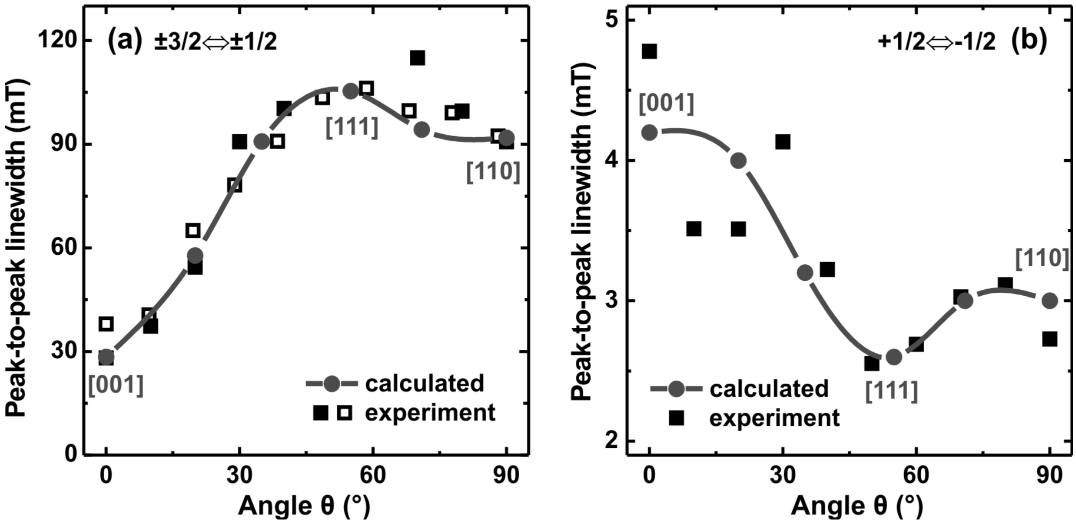


Figure 7.4: Calculated $\Delta m_j = 1$ EPR linewidths as a function of the magnetic field direction parametrized by the angle between the [001] direction (0°) and the [110] direction (90°). (a) We compare calculated linewidths for the $m_j = \pm 3/2 \Longleftrightarrow m_j = \pm 1/2$ transitions (filled circles) with new experimental data presented in Ref. [129] (filled squares) and previously reported data from Ref. [128] (open squares). (b) Same for the $m_j = +1/2 \Longleftrightarrow m_j = -1/2$ transition.

demonstrates that the line-broadening is in fact an isotopic effect. More details on these calculations together with an explanation of the fine structure can be found in Ref. [129].

We have repeated the calculation of the EPR line-broadening for different directions of the magnetic field. Fig. 7.4 shows the calculated peak-to-peak linewidths of the individual $\Delta m_j = 1$ transitions as a function of the magnetic field direction, ranging from [001] (0°) to [110] (90°). As can be deduced from the figure, the line broadening is strongly anisotropic. We find the linewidths of the $m_j = \pm 3/2 \Longleftrightarrow m_j = \pm 1/2$ transitions to have a minimum for $B \parallel [001]$ and a maximum for $B \parallel [111]$. For the $m_j = +1/2 \Longleftrightarrow m_j = -1/2$ transition, the situation gets reversed. Comparison with experimental data again shows excellent agreement. We note that the only free parameter used in these calculations is the potential Q at the central grid node that has been chosen to obtain the correct mixing of heavy and light hole states, which is responsible for the inhomogenous energy spacing between the Zeeman levels. Concretely, we have fitted the parameter only once to reproduce the experimental splitting of 49 mT between the transition energies $E\,(+3/2) - E\,(+1/2)$ and $E\,(1/2) - E\,(-1/2)$ [shown in Fig. 7.3(b)] for the magnetic field lying in the [001] direction.

Altogether, we have been able to successfully explain the anisotropic broadening of EPR signals for boron in natural silicon by isotopic shifts of the acceptor Zeeman levels. In addition, we would like to note that the present analysis forms an independent verification of the band offsets between the different silicon isotopes given in Ref. [130].

7.3 Donor wave functions in silicon nanocrystals

Phosphorus donors in bulk silicon can again be described in the hydrogenic impurity model, where the donor electron moves in the electrostatic potential of the singly charged donor ion screened by the bulk dielectric constant $\varepsilon_{\mathrm{bulk}} = 11.7$. The ground-state envelope function of the donor electron is shown schematically in Fig. 7.5. In silicon nanocrystals (Si-NCs), the electronic states are known to become more localized [133], leading to different ionization energies and doping properties compared to the bulk. However, it is not known which mechanisms dominate the electronic structure modifications. Existing models can be classified into the following two groups. In simple quantum confinement models, the donor wave function deviates strongly from the hydrogen-like function due to the influence of the confining potential V_0, as has been sketched in Fig. 7.5(a). On the other hand, it has been predicted in many theoretical investigations [134, 138, 139] that the screening of the Coulomb potential is reduced in nanocrystals. In a pure dielectric confinement model, the donor electron localization arises from a size-dependent dielectric screening $\varepsilon_{\mathrm{NC}}(R)$, which enters the Coulomb potential $V_c(r) = -1/[\varepsilon_{\mathrm{NC}}(R)\, r]$ between the impurity nucleus and the donor electron (R denotes the radius of the nanocrystal and r is the distance from the nucleus). This can also lead to an increase of the electron density at the impurity nucleus $|\Psi(0)|^2$, as shown schematically in Fig. 7.5(b).

Method

In order to clarify the relative contributions of dielectric confinement and quantum confinement to the localization of donor states in nanocrystals, we use an approach that considers both the surface confining potential and a size-dependent screening of the Coulomb potential. Concretely, we rely on the single-band effective-mass model introduced in Sec. 1.2.1 that has been augmented by a suitable potential. The Hamiltonian is given by

$$H(\mathbf{r}) = \frac{1}{2}\mathbf{p}^T \frac{1}{\hat{m}^*}\mathbf{p} + V(r), \tag{7.6}$$

where $\hat{m}^*$ denotes the effective 3×3 mass tensor for a silicon delta valley with longitudinal and transverse masses of $m_l = 0.92 m_0$ and $m_t = 0.19 m_0$, respectively. The impurity nucleus is represented by a positive charge at the center of a spherical nanocrystal. So, the potential energy of the donor electron is given by

$$V(r) = -\frac{e^2}{\varepsilon_{\mathrm{NC}}(R)\, r} + W(r) \text{ for } 0 < r \leq R, \tag{7.7}$$

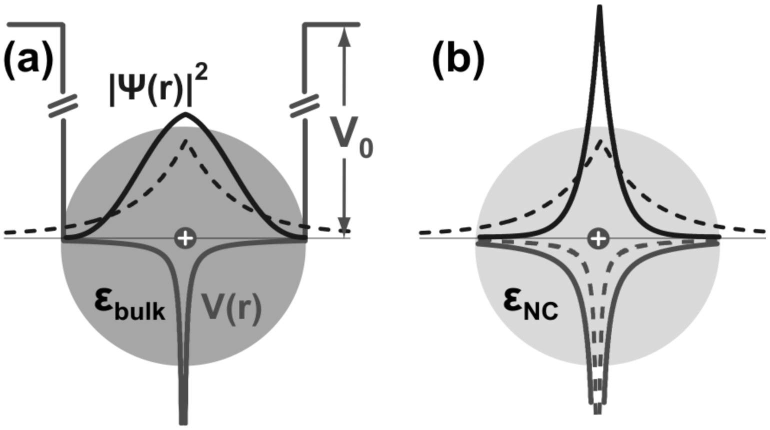


Figure 7.5: Schematic cross sections of the donor ground state in a semiconductor nanocrystal. (a) In the quantum confinement model, a large boundary barrier V_0 is added to the hydrogen-like potential $V(r)$ of a donor impurity in a bulk semiconductor. The solid black line sketches the deformation of the wave function with respect to the bulk case (dashed line). (b) In the dielectric confinement model, the bulk dielectric constant $\varepsilon_{\text{bulk}}$ is replaced by a smaller value ε_{NC}, which depends on the size of the nanocrystal. By this, the screening of the positive donor charge is reduced within the nanocrystal, leading to an increase of the probability density at the nucleus.

where we model the confining potential with a hard wall potential $V_0 \to +\infty$ for $r > R$. For the change of the dielectric screening, we take the relation

$$\varepsilon_{\text{NC}}(R) = 1 + \frac{\varepsilon_{\text{bulk}} - 1}{1 + (\alpha/R)^l}, \tag{7.8}$$

which has been derived from a generalization of the Penn model to nanocrystals [134]. Theoretical estimations of the parameters α and l include calculations of absorption spectra using empirical pseudopotentials [134] or ab-initio methods [138], both predicting a similar reduction of the electronic screening with respect to the bulk. In our calculations, we use the values $\alpha = 0.97$ nm and $l = 1.3$ determined in Ref. [138]. The use of slightly different values obtained by other authors gives nearly the same results and does not alter our conclusions. For our Si-NCs, we assume the same central cell correction parametrization as established previously for P in bulk Si [139]. I.e. for $W(r)$ in Eq. (7.7) we again use Eq. (7.3), but with reversed overall sign. Furthermore, the parameters α, β, and γ now equal 0.7572, 0.3123, and 2.044 of the reciprocal Bohr radius, respectively, and $A = 1.175$. In order to avoid the singularity at $r = 0$, the potentials are replaced by $-Q\delta_{r,0}$ for the central grid node, following Ref. [139]. Here, the parameter Q has been fitted to obtain the experimental binding energy of a P donor in bulk Si (45.6 meV) for the case of a very large Si-NC ($R = 25$ nm). This simple model has the advantage of being applicable also to large nanocrystals, unlike previous first principles approaches [140]. In addition, the individual contributions of dielectric confinement and quantum confinement can be easily switched on and off.

Using this model, we now calculate the electron ground state probability density as a function of the nanocrystal radius. The peak probability density at the donor nucleus $|\Psi(0)|^2$ can be probed in EPR experiments via the Fermi contact hyperfine splitting $A \propto |\Psi(0)|^2$ arising from the interaction between the electron spin and the spin I of the impurity nucleus. For ^{31}P $(I = 1/2)$ in bulk Si, the hyperfine splitting is $A_{\mathrm{bulk}} = 4.2$ mT [141]. The size dependent hyperfine splitting $A(R)$ is then calculated from the relative change of the electron probability density at the donor nucleus

$$A(R) = A_{\mathrm{bulk}} \frac{|\psi_R(0)|^2}{|\psi_\infty(0)|^2}. \tag{7.9}$$

Here, the wave function $\psi_\infty(0)$ representing the limit $R \to \infty$, was approximated by a Si-NC with a radius of 25 nm.

Results

In Fig. 7.6, we show our calculated results (solid line) that indicate a strong increase of the hyperfine interaction with the reduction of the nanocrystal radius R. In addition, we compare our results with experimental data (circles) and find excellent agreement. In order to elucidate the sole contributions of quantum confinement and reduced dielectric screening, we have considered two additional variations of the model. In the pure quantum confinement model (dotted line), the size-dependence of the dielectric screening is neglected by substituting $\varepsilon_{\mathrm{NC}}(R)$ with $\varepsilon_{\mathrm{bulk}}$ in Eq. (7.7) [cf. Fig. 7.5(a)]. In the pure dielectric confinement model (dashed line), the confining surface potential is neglected by setting $V_0 = 0$ [cf. Fig. 7.5(b)]. As can be deduced from Fig. 7.6, the pure quantum confinement model leads to a significant increase of the hyperfine splitting only for small radii below $R = 5$ nm. By contrast, the dielectric effect on the confinement results in a continuous increase already at large radii, but is less important in the limit of small R.

Experimentally, only very small nanocrystals with a radius between 2 and 3.5 nm have been investigated so far (open circles) [133]. Here, the increase of hyperfine splitting could be explained solely on the basis of quantum confinement due to the potential barrier V_0 of the surrounding medium. Namely, the dependence of the hyperfine coupling on the nanocrystal radius R has been described with a R^{-3} law [133], which corresponds to the localization of an electron in a spherical potential of infinite height without any Coulomb potential. This interpretation can be verified also by our more realistic quantum confinement model (dotted line). In order to clarify, whether this general conclusion also holds for larger nanocrystals, the group of Prof. Martin Brandt has performed measurements of hyperfine splittings associated with P donors for Si-NCs in the range of 3 to 16 nm (full circles), using electrically-detected magnetic resonance (EDMR) [135]. Although the two experimental sets of data agree

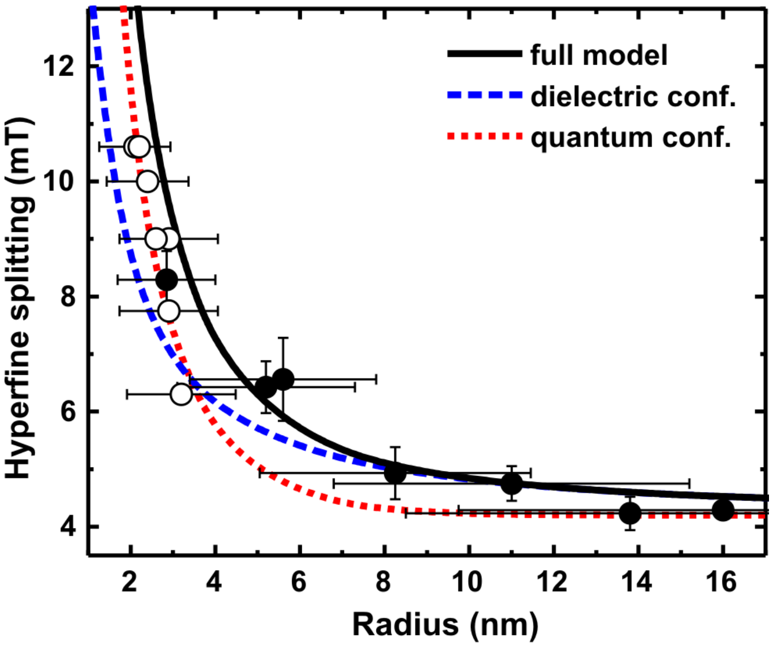


Figure 7.6: Calculated probability density of the donor ground state at the nucleus parametrized by the appropriate ^{31}P hyperfine splitting, as a function of the nanocrystal radius. We show results obtained by a pure quantum confinement model (dotted line), a pure dielectric confinement model (dashed line), and our full model where both effects are considered (solid line). Comparison with recent experimental data [135] (full circles) and previous data reported in Ref. [133] (open circles) shows excellent agreement.

reasonably, with a discrepancy at most in one single data point, the pure quantum confinement model used to describe the previous data, clearly deviates from the more recent results. On the other hand, the agreement with our full model (solid line) is striking and demonstrates that the reduction of dielectric screening is dominant for nanocrystals in the large size range, whereas for Si-NCs with radii below 6 nm both quantum confinement and dielectric confinement contribute to donor localization. By this, we have demonstrated that dielectric effects are very important in understanding nanoscale phenomena such as dopant confinement.

Supplementary information

Finally, we would like to note some further details that have been considered in the present calculations. As our model is based on a spherical Si-NC in vacuum, a realistic treatment of the problem in principle has to take into account the effect of self-polarization fields. They can be understood as the electric fields of image charges that are induced by the donor ion and donor electron due to the change of the dielectric constant at the Si-NC/vacuum boundary. Analytic expressions for the contributions of these effects to the Hamiltonian of the donor electron can be found in Eqs. 5 and 6 of Ref. [139]. While, in the case of the central donor position, the image charge of the donor ion only contributes a constant term to the Hamiltonian and cannot alter the donor wave function, the self-polarization term of the donor electron always has to be taken into account. Although, we find this self-polarization term to slightly increase the hyperfine splitting, it is by far not sufficient to explain the experimentally observed enhancement of the hyperfine coupling with respect to the pure quantum confinement model. For the sake of simplicity and due to the almost negligible magnitude of the polarization effect, we have omitted its discussion in the previous parts.

We should also note that the experiment does not have control over the location of P inside the Si-NCs. In order to estimate the influence of the donor position on the hyperfine splitting, we have also considered off-center donor positions. Here, we observe a small increase of the hyperfine splitting when the donor is moved from the center of the Si-NC toward the surface. A maximum is reached at approximately $3/4R$, followed by a rather sharp drop of the hyperfine coupling for positions even closer to the nanocrystal boundary. However, we believe that our model overestimates these variations due to its continuous medium character that does not accurately describe material interface regions. This assumption is supported by recent ab initio pseudopotential calculations [140]. Here, the amplitude of variation of the hyperfine splitting that is calculated when P is moved from the center toward the surface, is shown to decreases as the size of the nanocrystal increases, being only 10% for the smallest nanocrystals investigated experimentally ($R \sim 3$ nm). Thus, for a realistic distribution of donor positions in an ensemble of Si-NCs the resulting uncertainty of the hyperfine splitting is estimated to be far below the experimental error bars shown in Fig. 7.6.

7.4 Summary

In this chapter, we have investigated electronic structure properties of shallow impurities in silicon. Concretely, we have studied boron acceptors in bulk silicon subject to external magnetic fields. Here, we have calculated the isotopic shift of acceptor Zeeman levels via isotope induced local fluctuations of the band gap. We show that the inhomogenous broadening of EPR lines in natural silicon can be successfully explained solely by this effect [129].

In addition, we have investigated the contributions of quantum confinement and suppressed dielectric screening to the localization of donor wave functions in silicon nanocrystals. We observe a strong increase of the probability density at the position of the donor nucleus with the reduction of the nanocrystal radius. Our results are in excellent agreement with recent measurements [135] and show that dielectric effects, which have been neglected in the interpretation of most previous experiments, play a critical role.

Chapter 8

Summary and outlook

The main topic of this work has been the study of spin-related electronic structure properties of semiconductor nanostructures in magnetic fields. Here, the goal has been to propose a concrete application for solid state quantum information processing by numerical calculations within realistic three-dimensional models that include the detailed geometry and material composition. However, this study requires the solution of the Schrödinger equation for a mesoscopic semiconductor system, for which it is known that an accurate nonperturbative incorporation of the magnetic field requires special care to ensure gauge-invariant results.

Therefore, we have developed a novel general method to solve the multiband Schrödinger equation including strain, relativistic effects, and an external magnetic field in an efficient and manifestly gauge-invariant manner. It is based on the concept of gauge covariant derivatives and has been adapted from lattice gauge theory. We have successfully applied this method to investigate the quantitative electronic structure of realistic, experimentally realizable quantum dots subject to magnetic fields. From the magnetic energy spectrum, we have identified the gyromagnetic factors as the most application relevant properties, since these g factors characterize the splitting of spin states in magnetic fields and their spin precession frequency. We find them to be in fact anisotropic tensors and to strongly vary with the size, shape, and composition of the quantum dots. The limited bias tunability of g factors in single quantum dots has motivated us to extend the analysis to pairs of coupled dots. Indeed, we predict a giant electrically controllable anisotropy of hole g factors in self-assembled quantum dot molecules that allows ultrafast and coherent single-spin manipulations in a static magnetic field. This effect is used to propose a concrete realization of an efficient single spin-qubit gate. The accuracy of our calculations is verified by the fact that we are able to quantitatively explain experimentally observed resonant enhancements of exciton g tensor components for vertical magnetic fields without any fitting parameters. The experimental realization of our theoretical predictions is currently under investigation by the group of Prof. Jonathan Finley.

A second goal of this work has been the investigation of optoelectronic properties of heterostructures with a broken-gap band alignment. Here, we have faced the problem that standard multiband envelope function approaches fail to calculate the correct free-carrier charge density, because the separate occupation of electron and hole states is incompatible with a strong hybridization of conduction band and valence band states. Therefore, we have developed a novel charge self-consistent electronic structure scheme that remains in the electron framework throughout and occupies all included subbands according to Fermi statistics, before subtracting a positive background ionic charge to guarantee charge neutrality. With this procedure, we have calculated optical interband and intersubband transition energies of InAs/GaSb broken-gap superlattices. Such structures are of great interest for the fabrication of infrared lasers and infrared detectors. We obtain excellent agreement with experimental data and partly revise previous interpretations that lacked a consistent theoretical model.

With this work, we have also contributed to the field of silicon based quantum computing, where the quantum information is stored in the spins of shallow impurity states. In close collaboration with the experimental group of Prof. Martin Brandt, we have analyzed the influence of the isotope ratio in natural silicon on the spin energy levels of boron acceptors in external magnetic fields. We have successfully explained the broadening of EPR spin transition lines by isotope induced shifts of band gap energies. In addition, we have clarified that dielectric effects play a critical role on dopant confinement in silicon nanocrystals. This is crucial for understanding the process of electronic doping in nanostructures.

A final and very important benefit of the present work has been the development of the nanodevice simulation package next**nano**++. The software provides a global insight into a wide range of electronic, optical, and transport characteristics of mesoscopic semiconductor structures with virtually any geometry and combination of semiconducting materials. It is freely available and has many users in the scientific community worldwide. Due to its modular structure and the heavy usage of modern object oriented programming techniques, the software explicitly allows to easily extend its capabilities.

For the near future, it would be interesting to investigate the influence of exchange-correlation effects on the spin splittings of confined carriers and therefore the g factors. The magnetization of semiconductor nanostructures filled with a few charge carriers might also be relevant. In addition, an extension of the gauge-invariant discretization scheme to periodic structures would be useful. By this, one could also calculate magneto-optical transition rates for the quasi one-dimensional broken-gap heterostructures. This would allow more detailed comparison with experiments, which mostly rely on high in-plane magnetic fields to brake the selection rules. Another interesting future capability would be the calculation of carrier transport in magnetic fields. This would allow one to study, e.g., quantum Hall edge channels and realistic spintronic devices. However, this first requires the implementation of more sophisticated quantum transport models in next**nano**++, which are currently in development.

Appendix A

Wurtzite k·p Hamiltonian

In Sec. 1.2.1, we have presented the **k·p** Hamiltonian of the zincblende crystal structure. In the following, we will provide analogous relations for the wurtzite structure that are needed, e.g., to describe the group-III nitrides GaN, AlN, InN, and their ternary compounds.

For wurtzite materials, only the band extrema at the Γ-point ($\mathbf{k}_0 = 0$) are physically relevant, so that the single band EMA Hamiltonian (1.14) reduces to

$$\hat{H}^{\mathrm{EMA}}\left(\mathbf{k}\right) = E^{\Gamma} + \frac{\hbar^2}{2}\mathbf{k}^T \frac{1}{\hat{m}^*}\mathbf{k}, \tag{A.1}$$

with a diagonal effective mass tensor $\hat{m}^*$

$$\hat{m}^* = \left(m_c^* - m_a^*\right)\mathbf{e}_c\mathbf{e}_c^T + m_a^* 1^{3\times 3}, \tag{A.2}$$

where m_c^* and m_a^* are the effective masses along the hexagonal c-axis ($\mathbf{e}_c = [0001]$) and the perpendicular directions, respectively. In principle, this model can be used for electrons and holes. However, more accurate results can be obtained in the coupled eight-band model.

In the basis of spin-resolved zone-center conduction and valence band Bloch functions

$$\left|q\sigma\right\rangle \in \left\{\left|s\uparrow\right\rangle, \left|s\downarrow\right\rangle, \left|x_1\uparrow\right\rangle, \left|x_2\uparrow\right\rangle, \left|x_2\uparrow\right\rangle, \left|x_1\downarrow\right\rangle, \left|x_2\downarrow\right\rangle, \left|x_2\downarrow\right\rangle\right\}, \tag{A.3}$$

the eight-band Hamiltonian reads [142]

$$\hat{H}^{8\times 8}_{\mathrm{wz}}\left(\mathbf{k}\right) = \begin{pmatrix} \hat{H}^{2\times 2}_{\mathrm{wz,cc}}\left(\mathbf{k}\right) & \hat{H}^{2\times 6}_{\mathrm{wz,cv}}\left(\mathbf{k}\right) \\ \hat{H}^{6\times 2}_{\mathrm{wz,vc}}\left(\mathbf{k}\right) & \hat{H}^{6\times 6}_{\mathrm{wz,vv}}\left(\mathbf{k}\right) + \hat{H}^{6\times 6}_{\mathrm{wz,so}} + \hat{H}^{6\times 6}_{\mathrm{wz,cf}} \end{pmatrix} + \hat{H}^{8\times 8}_{\mathrm{wz},\varepsilon}, \tag{A.4}$$

with

$$\hat{H}^{2\times 2}_{\mathrm{wz,cc}}\left(\mathbf{k}\right) = \left(E_c + k_1 S_2 k_1 + k_2 S_2 k_2 + k_3 S_1 k_3\right) \otimes 1^{2\times 2}, \tag{A.5}$$

$$\hat{H}^{2\times 6}_{\mathrm{wz,cv}}\left(\mathbf{k}\right) = \begin{pmatrix} \mathrm{i}P_2 k_1 + k_2 B_1 k_3 & \mathrm{i}P_2 k_2 + k_3 B_2 k_1 & \mathrm{i}P_1 k_3 + k_1 B_3 k_2 \end{pmatrix} \otimes 1^{2\times 2}, \tag{A.6}$$

$$\hat{H}^{6\times 2}_{\mathrm{wz,vc}}\left(\mathbf{k}\right) = \left[\hat{H}^{2\times 6}_{\mathrm{wz,cv}}\left(\mathbf{k}\right)\right]^{\dagger}, \tag{A.7}$$

$$\hat{H}^{6\times 6}_{\mathrm{wz,vv}}(\mathbf{k}) = \left(E_v + \frac{\hbar^2}{2m_0}\sum_{i=1}^{3} k_i^2\right) \otimes 1^{6\times 6} + \hat{H}^{3\times 3}_{\mathrm{wz,vv}}(\mathbf{k}) \otimes 1^{2\times 2}, \tag{A.8}$$

$$\hat{H}^{3\times 3}_{\mathrm{wz,vv}}(\mathbf{k}) = \begin{pmatrix} k_1L_1k_1 + k_2M_1k_2 + k_3M_2k_3 & k_1N_1^+k_2 + k_2N_1^-k_1 & k_1N_2^+k_3 + k_3N_2^-k_1 \\ k_1N_1^-k_2 + k_2N_1^+k_1 & k_1M_1k_1 + k_2L_1k_2 + k_3M_2k_3 & k_2N_2^+k_3 + k_3N_2^-k_2 \\ k_1N_2^-k_3 + k_3N_2^+k_1 & k_2N_2^-k_3 + k_3N_2^+k_2 & k_1M_3k_1 + k_2M_3k_2 + k_3L_2k_3 \end{pmatrix} \tag{A.9}$$

Here, the parameters S_1 and S_2 characterize the conduction band effective masses, P_1 and P_2 are the interband coupling matrix elements, and B_1, B_2, B_3 are bulk inversion asymmetry parameters. The Dresselhaus parameters $L_1, L_2, M_1, M_2, M_3, N_1, N_2$ can be written in terms of the more widely known Rashba parameters $A_1, ..., A_6$ [143]

$$L_1 = \frac{\hbar}{2m_0}\left(A_5 + A_4 + A_2 - 1\right),\ L_2 = \frac{\hbar}{2m_0}\left(A_1 - 1\right), \tag{A.10}$$

$$M_1 = \frac{\hbar}{2m_0}\left(A_4 + A_2 - A_5 - 1\right),\ M_2 = \frac{\hbar}{2m_0}\left(A_1 + A_3 - 1\right), \tag{A.11}$$

$$M_3 = \frac{\hbar}{2m_0}\left(A_2 - 1\right),\ N_1 = \frac{\hbar}{2m_0}2A_5,\ N_2 = \frac{\hbar}{2m_0}\sqrt{2}A_6. \tag{A.12}$$

The ordering of material parameters with respect to momentum operators can be derived using Burt's exact envelope function theory [15]. Following Ref. [144], the parameters $N_i^{\pm}$ are given in terms of tabulated ones

$$N_i^+ = N_i - M_i,\ N_i^- = M_i.\ (i = 1, 2) \tag{A.13}$$

The spin-orbit coupling of the valence bands is included via the term

$$\hat{H}^{6\times 6}_{\mathrm{wz,so}} = \begin{pmatrix} 0 & -i\Delta_2 & 0 & 0 & 0 & \Delta_3 \\ i\Delta_2 & 0 & 0 & 0 & 0 & -i\Delta_3 \\ 0 & 0 & 0 & -\Delta_3 & i\Delta_3 & 0 \\ 0 & 0 & -\Delta_3 & 0 & i\Delta_2 & 0 \\ 0 & 0 & -i\Delta_3 & -i\Delta_2 & 0 & 0 \\ \Delta_3 & i\Delta_3 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{A.14}$$

with $\Delta_2 = \Delta_{\mathrm{so}}^{(c)}/3$ and $\Delta_3 = \Delta_{\mathrm{so}}^{(a)}/3$. Wurtzite materials are subject to an additional crystal field splitting

$$\hat{H}^{6\times 6}_{\mathrm{wz,cf}} = \begin{pmatrix} \Delta_1 & 0 & 0 \\ 0 & \Delta_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \otimes 1^{2\times 2}, \tag{A.15}$$

that lifts the degeneracy of the heavy hole band and the light hole band for $\mathbf{k} = 0$, already in the absence of strain.

In the eight-band Hamiltonian (A.4), the strain contribution via deformation potentials is given by [142]

$$\hat{H}^{8\times 8}_{\mathrm{wz},\varepsilon} = \begin{pmatrix} H_{\mathrm{wz},\varepsilon} \otimes 1^{2\times 2} & 0 \\ 0 & \hat{H}^{3\times 3}_{\mathrm{wz},\varepsilon} \otimes 1^{2\times 2} \end{pmatrix}. \tag{A.16}$$

Here, the strain induced energy shift of the conduction bands is given by

$$H_{\mathrm{wz},\varepsilon} = \Xi_t \left(\varepsilon_{11} + \varepsilon_{22}\right) + \Xi_l \varepsilon_{33}, \tag{A.17}$$

where Ξ_l and Ξ_t are the longitudinal (with respect to the hexagonal c-axis) and the transversal deformation potential, respectively. The valence-band part reads

$$\hat{H}^{3\times3}_{\mathrm{wz},\varepsilon} = \begin{pmatrix} l_1\varepsilon_{11} + m_1\varepsilon_{22} + m_2\varepsilon_{33} & n_1\varepsilon_{12} & n_2\varepsilon_{13} \\ n_1\varepsilon_{12} & l_1\varepsilon_{22} + m_1\varepsilon_{11} + m_2\varepsilon_{33} & n_2\varepsilon_{23} \\ n_2\varepsilon_{13} & n_2\varepsilon_{23} & l_2\varepsilon_{33} + m_3\left(\varepsilon_{11} + \varepsilon_{22}\right) \end{pmatrix}. \tag{A.18}$$

Here, the parameters $l_1, l_2, m_1, m_2, m_3, n_1, n_2$ depend on the deformation potentials $d_1, ..., d_6$ [143] via the relations

$$\begin{aligned} l_1 &= d_5 + d_4 + d_2,\ l_2 = d_1, \\ m_1 &= d_4 + d_2 - d_5,\ m_2 = d_1 + d_3,\ m_3 = d_2, \\ n_1 &= 2d_5,\ n_2 = \sqrt{2}d_6. \end{aligned} \tag{A.19}$$

Appendix B

Local spin density approximation (LSDA)

In this appendix, we present the parametrization of the exchange-correlation potentials introduced in Sec. 1.3.3. We have implemented the local spin density method by invoking the LSDA functional of Perdew and Zunger [38]. This method accounts for the spin polarization

$$\zeta(\mathbf{x}) = \frac{n_\uparrow(\mathbf{x}) - n_\downarrow(\mathbf{x})}{n(\mathbf{x})}, \tag{B.1}$$

with $n(\mathbf{x}) = n_\uparrow(\mathbf{x}) + n_\downarrow(\mathbf{x})$,where $n_\uparrow(\mathbf{x})$ and $n_\downarrow(\mathbf{x})$ denote the spin-up and spin-down carrier densities, respectively. The LSDA exchange-correlation potential is given by

$$\begin{aligned} V_{xc}^{\uparrow\downarrow}(\mathbf{x}) &= \frac{\partial}{\partial n}\left[n\varepsilon_{xc}(n)\right]\Bigg|_{n=n_{\uparrow\downarrow}(\mathbf{x})} \\ &= V_x^{\uparrow\downarrow}(\mathbf{x}) + V_c^{\uparrow\downarrow}(\mathbf{x}), \end{aligned} \tag{B.2}$$

with the exchange (x) and correlation (c) energies

$$\varepsilon_{xc}(\mathbf{x}) = \varepsilon_x(\mathbf{x}) + \varepsilon_c(\mathbf{x}), \tag{B.3}$$

$$\varepsilon_{x/c}(\mathbf{x}) = \left\{\varepsilon_{x/c}^U(r_s\,(\mathbf{x})) + \left[\varepsilon_{x/c}^P(r_s\,(\mathbf{x})) - \varepsilon_{x/c}^U(r_s\,(\mathbf{x}))\right] f(\zeta\,(\mathbf{x}))\right\} H^*, \tag{B.4}$$

and appropriate potentials

$$\begin{aligned} V_{x/c}^{\uparrow\downarrow}(\mathbf{x}) = \big\{ & V_{x/c}^U(r_s(\mathbf{x})) + \left[V_{x/c}^P(r_s(\mathbf{x})) - V_{x/c}^U(r_s(\mathbf{x}))\right] f(\zeta(\mathbf{x})) \\ & + \left[\varepsilon_{x/c}^P(r_s(\mathbf{x})) - \varepsilon_{x/c}^U(r_s(\mathbf{x}))\right] (\pm 1 - \zeta(\mathbf{x}))\, f'(\zeta(\mathbf{x}))\big\} H^*. \end{aligned} \tag{B.5}$$

Here, energies and potentials are interpolated between the LDA expressions for a homogenous carrier gas that is fully polarized (P: $\zeta = 1$) or completely unpolarized (U: $\zeta = 0$) using the interpolation formulas

$$f(\zeta) = \frac{1}{2(2^{1/3} - 1)}[(1+\zeta)^{4/3} + (1-\zeta)^{4/3} - 2], \tag{B.6}$$

$$f'(\zeta) = \frac{1}{2^{1/3} - 1}\frac{2}{3}[(1+\zeta)^{1/3} - (1-\zeta)^{1/3}]. \tag{B.7}$$

Furthermore, we introduced the characteristic length scale

$$r_s(\mathbf{x}) = \frac{1}{a_B^*}\left(\frac{3}{4\pi n(\mathbf{x})}\right)^{1/3}. \tag{B.8}$$

Energies are generally measured in units of effective Hartrees $H^* = m^*/\varepsilon^2 H$, and lengths in units of effective Bohr radii $a_B^* = \varepsilon/m^* a_B$, where ε is the (dimensionless) static dielectric constant and m^* is the (dimensionless) effective mass. We use the static rather than the electronic (also referred to as "high frequency") dielectric constant since we assume that, in the majority of applications, the kinetic energy of the electrons will not exceed the optical phonon energy of typically 50 meV so that the phonons are able to screen the free carriers. One may argue that it is more appropriate to use the electronic dielectric constant in the eight-band case, but we use the same value throughout for consistency reasons. The value to be used for the effective mass depends on the **k·p** band model as shown in the table below. Note that the dominating exchange potential is actually independent of the effective mass so we are not introducing a noticeable inconsistency between single-band versus eight-band solutions in this way. For an ellipsoidal valley, we use the same average density of state mass, irrespective of the actual density of states of a particular valley. The rationale behind this is the fact that the usage of a different density functional for each valley is in conflict with the variational Hohenberg-Kohn principle.

Single-band situation (spherical)	m^*
Single-band situation (ellipsoidal)	$m^* = g^{2/3}\left(m_l m_t^2\right)^{1/3}$
Six-band situation:	$m^* = 1$
Eight-band situation	$m^* = 1$

(B.9)

Here, g is a valley degeneracy factor; it counts the number of valleys that remain equivalent independent of the symmetry and strain. This factor is 2 only in Si where a maximum of 3 (out of 6) valleys can become nonequivalent; for all other cases with valleys at the Brillouin zone boundary it is 1.

By inserting the concrete expressions for the polarized and unpolarized cases, the exchange energies and potentials can be simplified to

$$\varepsilon_x(\mathbf{x}) = -\frac{3}{4}\left(\frac{9}{4\pi^2}\right)^{1/3}\frac{1}{r_s(\mathbf{x})}\frac{1}{2}[(1+\zeta(\mathbf{x}))^{4/3} + (1-\zeta(\mathbf{x}))^{4/3}]H^*, \tag{B.10}$$

$$V_x^{\uparrow\downarrow}(\mathbf{x}) = -\left(\frac{9}{4\pi^2}\right)^{1/3}\frac{1}{r_s(\mathbf{x})}\left[1 \pm \zeta(\mathbf{x})\right]^{1/3} H^*. \tag{B.11}$$

For the correlation energies and potentials, we use the parametrization of Ref. [38]

$$\varepsilon_c^i(r_s) = \left\{\begin{array}{cc} \gamma_i/\left(1+\beta_1^i\sqrt{r_s}+\beta_2^i r_s\right) & r_s \geq 1 \\ (A_i \ln r_s + B_i + C_i r_s \ln r_s + D_i r_s & r_s < 1 \end{array}\right., \tag{B.12}$$

$$V_c^i(r_s) = \left\{\begin{array}{cc} \varepsilon_c^i\left(1+\frac{7}{6}\beta_1^i\sqrt{r_s}+\frac{4}{3}\beta_2^i r_s\right)/\left(1+\beta_1^i\sqrt{r_s}+\beta_2^i r_s\right) & r_s \geq 1 \\ A_i \ln r_s + \left(B_i - \frac{1}{3}A_i\right) + \frac{2}{3}C_i r_s \ln r_s + \frac{1}{3}\left(2D_i - C_i\right) r_s & r_s < 1 \end{array}\right., \tag{B.13}$$

with $i = U, P$ and the set of parameters given in the following table.

i	U	P
γ	-0.1423	-0.0843
β_1	1.0529	1.3981
β_2	0.3334	0.2611
A	0.0311	0.01555
B	-0.048	-0.0269
C	0.0020	0.0007
D	-0.0116	-0.0048

(B.14)

Appendix C

Analytic expressions for strain in heterostructures

In Sec. 1.4, the strain in multi-dimensional nanostructures has been calculated by minimizing the total elastic energy via the differential equation (1.83). In quasi one-dimensional structures, the range of possible geometries is strongly reduced to heterostructures only. In this situation, analytic expressions can be found for the strain in layers that exhibit a lattice mismatch with respect to the substrate. For an arbitrary growth direction $[hkl]$, the following general relation can be found for the distortion tensor in the Cartesian basis [42, 145, 146]

$$\hat{e} = \begin{pmatrix} u_0 & 0 & 0 \\ 0 & u_0 & 0 \\ 0 & 0 & v_0 \end{pmatrix} + D \begin{pmatrix} n_1^2 D_1 & n_1 n_2 D_1 & n_1 n_3 D_1 \\ n_1 n_2 D_2 & n_2^2 D_2 & n_2 n_3 D_2 \\ n_1 n_3 D_3 & n_2 n_3 D_3 & n_3^2 D_3 \end{pmatrix}, \tag{C.1}$$

with the normalized vector

$$\mathbf{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \frac{1}{\sqrt{h^2 + k^2 + l^2}} \begin{pmatrix} h \\ k \\ l \end{pmatrix}. \tag{C.2}$$

In zincblende materials, the remaining variables are given by

$$u_0 = v_0 = \frac{a_s - a}{a}, \tag{C.3}$$

$$D_1 = D_y D_z,\ D_2 = D_x D_z,\ D_3 = D_x D_y, \tag{C.4}$$

$$\begin{aligned} D_x &= (C_{11} - C_{12} - C_{44})\, n_1^2 + C_{44}\left(n_2^2 + n_3^2\right), \\ D_y &= (C_{11} - C_{12} - C_{44})\, n_2^2 + C_{44}\left(n_1^2 + n_3^2\right), \\ D_z &= (C_{11} - C_{12} - C_{44})\, n_3^2 + C_{44}\left(n_1^2 + n_2^2\right), \end{aligned} \tag{C.5}$$

$$D = -u_0 \frac{C_{11} + 2C_{12}}{D_x D_y D_z + (C_{12} + C_{44})\left(n_1^2 D_1 + n_2^2 D_2 + n_3^2 D_3\right)}. \tag{C.6}$$

Here, a and a_s denote the lattice constants of the strained layer and the substrate, respectively, and C_{ij} are the elements of the fourth rank elasticity tensor in Voigt notation. For wurtzite crystal structure, the analogous relations read

$$u_0 = \frac{a_s - a}{a},\ v_0 = \frac{c_s - c}{c}, \tag{C.7}$$

$$\begin{aligned} D_1 = D_2 &= W_1\left(n_1^2 + n_2^2\right) + \frac{C_{33}W_1 - (C_{13} + C_{44})\,W_2}{C_{44}} n_3^2, \\ D_3 &= W_2 n_3^2 + \frac{C_{11}W_2 - (C_{13} + C_{44})\,W_1}{C_{44}}\left(n_1^2 + n_2^2\right), \end{aligned} \tag{C.8}$$

$$\begin{aligned} W_1 &= -u_0\left(C_{11} + C_{12}\right) - v_0 C_{13}, \\ W_2 &= -u_0 2C_{13} - v_0 C_{33}, \end{aligned} \tag{C.9}$$

$$D = \left[C_{11}\left(n_1^2 + n_2^2\right)^2 + C_{33}n_3^4 + \frac{C_{11}C_{33} - 2C_{13}C_{44} - C_{13}^2}{C_{44}} n_3^2\left(n_1^2 + n_2^2\right)\right]^{-1}. \tag{C.10}$$

Here, additional lattice constants c and c_s have been introduced for the hexagonal c-axis. The strain tensor $\hat{\varepsilon}$ can be obtained straightforwardly by symmetrizing Eq. (C.1).

Appendix D

Calculation of discrete approximate derivatives

In order to calculate semiconductor nanostructures, we have to solve several differential equations that have been introduced in chapter 1. In chapter 2, we have already discussed the numerical realization of this task. However, we often need to examine not only the solutions f of the differential equations but also derivatives $\partial_j f$ of these properties. Concretely, we are interested in the electric field $\mathbf{F}$ [Eq. (1.67)], the distortion tensor $\hat{e}$ [Eq. (1.70)], and the current density $\mathbf{j}$ [Eq. (1.106)] that have to be calculated in a post-processing step from the electrostatic potential ϕ [determined by Eq. (1.65)], the displacement vector $\mathbf{u}$ [Eq. (1.83)], and the quasi Fermi level E_F [Eq. (1.105)], respectively. Since these functions have been discretized for numerical solution, their derivatives can as well only be evaluated via discrete approximations.

After discretization and numerical solution of a Laplacian operator $\partial_i a\left(\mathbf{x}\right) \partial_j f\left(\mathbf{x}\right)$ [Eqs. (1.65), (1.83), and (1.105)], we have obtained a discrete vector $\mathbf{f}$ that approximates the continuous function f and is defined only on the N grid nodes $\mathbf{m}$ of the rectangular lattice introduced in Sec. 2.3. Since we have employed box integration discretization, the discrete approximate derivatives ($\partial_j f \to \delta_j f$) are defined on the perpendicular boundary segments of the control boxes. We now exemplify this issue for the two-dimensional grid shown in Fig. D.1 and consider a particular boundary segment $S\left(m_1 + 1/2, m_2\right)$ that belongs to the control box of the grid node $\left(m_1, m_2\right)$. Since $a\left(\mathbf{x}\right)$ is constant within each of the quadrants of the control box, while f is defined on grid nodes only, we split the line S into two parts $S\left(m_1 + 1/2, m_2^{\pm}\right)$ (in 3D we have four surface elements), where the discrete derivative is given by

$$\left(a\delta_1 f\right)\left(m_1 + 1/2, m_2^{\pm}\right) = a\left(m_1 + 1/2, m_2 \pm 1/2\right) \frac{f\left(m_1 + 1, m_2\right) - f\left(m_1, m_2\right)}{\varepsilon_1\left(m_1, +1\right)}. \tag{D.1}$$

In most situations, we are interested only in the average derivative in each lattice area. For the area defined by the center point $\left(m_1 + 1/2, m_2 - 1/2\right)$ this requires to sum the contributions from the surface elements $S\left(m_1 + 1/2, m_2^{-}\right)$ and

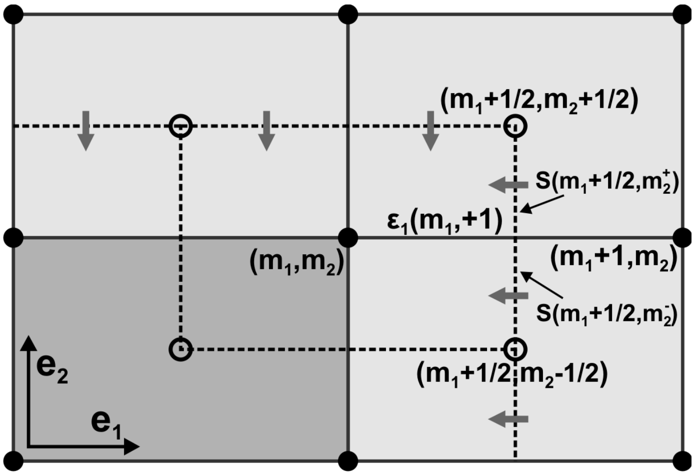


Figure D.1: Sketch of two-dimensional rectangular grid at the interface between a contact (dark gray area) and the inner simulation domain (light gray area). The current density is defined on the boundary segments of the control boxes (dashed lines). In order to calculate the current that flows into the contact (gray arrows), one has to integrate the current density over the surrounding control box surface.

$S\left(m_1+1/2,(m_2-1)^+\right)$, leading to the formula

$$\begin{aligned}(a\delta_1 f)\left(m_1+1/2,m_2-1/2\right) = \frac{a\left(m_1+1/2,m_2-1/2\right)}{2\varepsilon_1\left(m_1,+1\right)} &\left[f\left(m_1+1,m_2\right)\right. \\ &\left.+f\left(m_1+1,m_2-1\right)-f\left(m_1,m_2\right)-f\left(m_1,m_2-1\right)\right].\end{aligned} \tag{D.2}$$

For the d-dimensional case, we have to sum over 2^{d-1} $(d-1)$-dimensional surface elements, and the relation can be generalized to the form

$$(a\delta_j f)\left(\mathbf{m},\sigma/2\right) = \frac{a\left(\mathbf{m},\sigma/2\right)}{2^{d-1}\varepsilon_j\left(m_j,1\right)}\sum_{\sigma'}\left(2\sigma'_j-1\right)f\left(\mathbf{m},\sigma'\right), \tag{D.3}$$

with $\sigma_i = 1$ and $\sigma'_i \in \{0,1\}$ for $i = 1,...,d$. In the particular situation of the current density, we sometimes need to consider the individual contributions of the control box surface elements, separately. This happens at edges of contacts, when the current density is integrated over the contact surfaces in order to determine the total current flowing in or out of the contacts. In Fig. D.1, we have indicated the current contributions by gray arrows. As one can deduce from the figure, the corner lattice area associated with the center point $(m_1+1/2,m_2+1/2)$ needs to be treated differently from the other areas.

Appendix E

Continuum limit of gauge covariant discretization

In Sec. 3.3.3, we have shown that the correct continuum limit holds for the approximations of the first-order gauge covariant derivatives. In order to prove Eq. (3.37) for the second-order derivatives with $i = j$, we expand the connection U in Eq. (3.30) up to second order of ε

$$U\left(\mathbf{m}, \mathbf{m} + s_i\boldsymbol{\varepsilon}_i, i\right) = 1 + s_i \mathrm{i}\frac{e}{\hbar} A_i\left(\mathbf{m}\right)\varepsilon + s_i^2 \frac{1}{2}\mathrm{i}\frac{e}{\hbar}\left(\partial_i A_i\right)\left(\mathbf{m}\right)\varepsilon^2 - s_i^2 \frac{1}{2}\frac{e^2}{\hbar^2} A_i^2\left(\mathbf{m}\right)\varepsilon^2 + \mathcal{O}\left(\varepsilon^3\right). \tag{E.1}$$

For the wave functions $\mathbf{F}$, we have

$$\mathbf{F}\left(\mathbf{m} + s_i\boldsymbol{\varepsilon}_i\right) = \mathbf{F}\left(\mathbf{m}\right) + s_i\partial_i\mathbf{F}\left(\mathbf{m}\right)\varepsilon + s_i^2\frac{1}{2}\partial_i^2\mathbf{F}\left(\mathbf{m}\right)\varepsilon^2 + \mathcal{O}\left(\varepsilon^3\right). \tag{E.2}$$

By inserting Eqs. (E.1),(E.2) into Eq. (3.18), we get

$$\begin{aligned}\Delta_{ii}\mathbf{F}\left(\mathbf{m}\right) = \frac{1}{\varepsilon^2}\sum_{s_i} C_{ii}\left(s_i\right)\Big\{&\mathbf{F}\left(\mathbf{m}\right) + s_i\partial_i\mathbf{F}\left(\mathbf{m}\right)\varepsilon + s_i^2\frac{1}{2}\partial_i^2\mathbf{F}\left(\mathbf{m}\right)\varepsilon^2 \\ &+ s_i\,\mathrm{i}\frac{e}{\hbar}A_i\left(\mathbf{m}\right)\varepsilon\left[\mathbf{F}\left(\mathbf{m}\right) + s_i\partial_i\mathbf{F}\left(\mathbf{m}\right)\varepsilon\right] \\ &+ s_i^2\frac{1}{2}\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_i A_i\right)\left(\mathbf{m}\right) - \frac{e^2}{\hbar^2}A_i^2\left(\mathbf{m}\right)\right]\mathbf{F}\left(\mathbf{m}\right)\varepsilon^2\Big\} + \mathcal{O}\left(\varepsilon\right).\end{aligned} \tag{E.3}$$

Here the sum runs over $s_i \in \{0, \pm 1\}$. The first line in Eq. (E.3) must tend to $\partial_i^2\mathbf{F}\left(\mathbf{m}\right)$ as in the field-free case, which requires the coefficients to obey

$$\sum_{s_i} C_{ii}\left(s_i\right) = \sum_{s_i} C_{ii}\left(s_i\right)s_i = 0,\ \sum_{s_i} C_{ii}\left(s_i\right)s_i^2 = 2. \tag{E.4}$$

Using these relations, the entire expression Eq. (E.3) tends toward the correct continuum limit Eq. (3.37)

$$\Delta_{ii}\mathbf{F}\left(\mathbf{m}\right) = \left[\partial_i + \mathrm{i}\frac{e}{\hbar}A_i\left(\mathbf{m}\right)\right]^2\mathbf{F}\left(\mathbf{m}\right) + \mathcal{O}\left(\varepsilon\right). \tag{E.5}$$

For the mixed second-order derivatives, the diagonally adjacent neighbors are needed. For this case, we expand Eq. (3.31) analogous to Eq. (E.1)

$$\begin{aligned} U\left(\mathbf{m},\mathbf{m}+s_i\boldsymbol{\varepsilon}_i+s_j\boldsymbol{\varepsilon}_j,ij\right) = 1 &+ \left[s_i\mathrm{i}\frac{e}{\hbar}A_i\left(\mathbf{m}\right)+s_j\mathrm{i}\frac{e}{\hbar}A_j\left(\mathbf{m}\right)\right]\varepsilon \\ &+ s_i^2\frac{1}{2}\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_iA_i\right)\left(\mathbf{m}\right)-\frac{e^2}{\hbar^2}A_i^2\left(\mathbf{m}\right)\right]\varepsilon^2 \\ &+ s_is_j\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_iA_j\right)\left(\mathbf{m}\right)-\frac{e^2}{\hbar^2}A_i\left(\mathbf{m}\right)A_j\left(\mathbf{m}\right)\right]\varepsilon^2 \\ &+ s_j^2\frac{1}{2}\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_jA_j\right)\left(\mathbf{m}\right)-\frac{e^2}{\hbar^2}A_j^2\left(\mathbf{m}\right)\right]\varepsilon^2+\mathcal{O}\left(\varepsilon^3\right). \end{aligned} \tag{E.6}$$

Furthermore, we need

$$\begin{aligned} \mathbf{F}\left(\mathbf{m}+s_i\boldsymbol{\varepsilon}_i+s_j\boldsymbol{\varepsilon}_j\right) &= \mathbf{F}\left(\mathbf{m}\right)+\left[s_i\partial_i\mathbf{F}\left(\mathbf{m}\right)+s_j\partial_j\mathbf{F}\left(\mathbf{m}\right)\right]\varepsilon \\ &+ \left[s_i^2\frac{1}{2}\partial_i^2\mathbf{F}\left(\mathbf{m}\right)+s_is_j\partial_i\partial_j\mathbf{F}\left(\mathbf{m}\right)+s_j^2\frac{1}{2}\partial_j^2\mathbf{F}\left(\mathbf{m}\right)\right]\varepsilon^2+\mathcal{O}\left(\varepsilon^3\right). \end{aligned} \tag{E.7}$$

Inserting these relations into Eq. (3.18), gives ($s_i, s_j \in \{0, \pm1\}$)

$$\begin{aligned} \Delta_{ij}\mathbf{F}\left(\mathbf{m}\right) = \frac{1}{\varepsilon^2}\sum_{s_i,s_j}C_{ij}\left(s_i,s_j\right)\Big\{&\mathbf{F}\left(\mathbf{m}\right)+\left[s_i\partial_i\mathbf{F}\left(\mathbf{m}\right)+s_j\partial_j\mathbf{F}\left(\mathbf{m}\right)\right]\varepsilon \\ &+ \left[s_i^2\frac{1}{2}\partial_i^2\mathbf{F}\left(\mathbf{m}\right)+s_is_j\partial_i\partial_j\mathbf{F}\left(\mathbf{m}\right)+s_j^2\frac{1}{2}\partial_j^2\mathbf{F}\left(\mathbf{m}\right)\right]\varepsilon^2 \\ &+ \left[s_i\mathrm{i}\frac{e}{\hbar}A_i\left(\mathbf{m}\right)+s_j\mathrm{i}\frac{e}{\hbar}A_j\left(\mathbf{m}\right)\right]\left[\mathbf{F}\left(\mathbf{m}\right)+s_i\partial_i\mathbf{F}\left(\mathbf{m}\right)\varepsilon+s_j\partial_j\mathbf{F}\left(\mathbf{m}\right)\varepsilon\right]\varepsilon \\ &+s_i^2\frac{1}{2}\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_iA_i\right)\left(\mathbf{m}\right)-\frac{e^2}{\hbar^2}A_i^2\left(\mathbf{m}\right)\right]\mathbf{F}\left(\mathbf{m}\right)\varepsilon^2 \\ &+s_is_j\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_iA_j\right)\left(\mathbf{m}\right)-\frac{e^2}{\hbar^2}A_i\left(\mathbf{m}\right)A_j\left(\mathbf{m}\right)\right]\mathbf{F}\left(\mathbf{m}\right)\varepsilon^2 \\ &+s_j^2\frac{1}{2}\left[\mathrm{i}\frac{e}{\hbar}\left(\partial_jA_j\right)\left(\mathbf{m}\right)-\frac{e^2}{\hbar^2}A_j^2\left(\mathbf{m}\right)\right]\mathbf{F}\left(\mathbf{m}\right)\varepsilon^2\Big\}+\mathcal{O}\left(\varepsilon\right). \end{aligned} \tag{E.8}$$

Again, the field free case constricts the coefficients to

$$\sum_{s_i,s_j}C_{ij}\left(s_i,s_j\right)=\sum_{s_i,s_j}C_{ij}\left(s_i,s_j\right)s_i=0,\ \sum_{s_i,s_j}C_{ij}\left(s_i,s_j\right)s_j=0, \tag{E.9}$$

$$\sum_{s_i,s_j}C_{ij}\left(s_i,s_j\right)s_j^2=\sum_{s_i,s_j}C_{ij}\left(s_i,s_j\right)s_i^2=0,\ \sum_{s_i,s_j}C_{ij}\left(s_i,s_j\right)s_is_j=2. \tag{E.10}$$

Inserting these relations into Eq. (E.8), results in Eq. (3.37)

$$\Delta_{ij}\mathbf{F}\left(\mathbf{m}\right)=\left[\partial_i+\mathrm{i}\frac{e}{\hbar}A_i\left(\mathbf{m}\right)\right]\left[\partial_j+\mathrm{i}\frac{e}{\hbar}A_j\left(\mathbf{m}\right)\right]\mathbf{F}\left(\mathbf{m}\right)+\mathcal{O}\left(\varepsilon\right), \tag{E.11}$$

which has the correct limit.

List of Publications

- *Self-consistent multiband envelope function approach for InAs/GaSb type-II broken-gap heterostructures*, T. Andlauer and P. Vogl, in preparation for Phys. Rev. B.

- *Electron paramagnetic resonance of boron acceptors in silicon revisited*, A. R. Stegner, T. Andlauer, H. Tezuka, M. Stutzmann, K. M. Itoh, and M. S. Brandt, in preparation for Phys. Rev. B.

- *Dielectric screening versus quantum confinement of P-donors in silicon nanocrystals investigated by magnetic resonance*, R. N. Pereira, A. R. Stegner, T. Andlauer, K. Klein, M. S. Brandt, M. Stutzmann, and H. Wiggers, submitted to Phys. Rev. Lett.

- *Electrically controllable g tensors in quantum dot molecules*, T. Andlauer and P. Vogl, Phys. Rev. B. **79**, 045307 (2009).

- *Self-consistent electronic structure method for broken-gap superlattices*, T. Andlauer, T. Zibold, and P. Vogl, Proc. SPIE **7222**, 722211 (2009).

- *Gauge-invariant discretization in multiband envelope function theory and g factors in nanowire dots*, T. Andlauer, R. Morschl, and P. Vogl, Phys. Rev. B **78**, 075317 (2008).

- *nextnano: General Purpose 3D Simulations*, S. Birner, T. Zibold, T. Andlauer, T. Kubis, A. Trellakis, and P. Vogl, IEEE Trans. Electron Devices **54**, 2137 (2007).

- *The 3D nanometer device project nextnano: Concepts, Methods, Results*, A. Trellakis, T. Zibold, T. Andlauer, S. Birner, R. K. Smith, R. Morschl, P. Vogl, J. Comput. Electron. **5**, 285 (2006).

- *Efficient Solution of the Schrödinger-Poisson Equations in Semiconductor Device Simulations*, A. Trellakis, T. Andlauer, and P. Vogl, LSSC, Lecture Notes in Computer Science **3743**, 602 (2006).

- *Modeling of semiconductor nanostructures with nextnano*^{3}, S. Birner, S. Hackenbuchner, M. Sabathil, G. Zandler, J. A. Majewski, T. Andlauer, T. Zibold, R. Morschl, A. Trellakis, and P. Vogl, Acta Physica Polonica A **110**, 111 (2006).

Bibliography

[1] D. Loss and D. P. DiVincenzo, Phys. Rev. A **57**, 120 (1998).

[2] S. Datta and B. Das, Appl. Phys. Lett. **56**, 665 (1990).

[3] H. Ohno, L. Esaki, and E. E. Mendez, Appl. Phys. Lett. **60**, 3153 (1992).

[4] See http://www.wsi.tum.de/nextnano for obtaining the nextnano executables and related publications.

[5] S. Hackenbuchner, *Elektronische Struktur von Halbleiter-Nanobauelementen im thermodynamischen Nichtgleichgewicht*, vol. 48 of *Selected Topics of Semiconductor Physics and Technology* (Verein zur Förderung des Walter Schottky Instituts, Garching, 2002).

[6] T. Zibold, *Semiconductor based quantum information devices: Theory and simulations*, vol. 87 of *Selected Topics of Semiconductor Physics and Technology* (Verein zur Förderung des Walter Schottky Instituts, Garching, 2007).

[7] M. Sabathil, *Opto-electronic and quantum transport properties of semiconductor nanostructures*, vol. 67 of *Selected Topics of Semiconductor Physics and Technology* (Verein zur Förderung des Walter Schottky Instituts, Garching, 2004).

[8] E. L. Ivchenko and G. E. Pikus, *Superlattices and Other Heterostructures: Symmetry and Optical Phenomena*, vol. 110 of *Springer Series in Solid State Sciences* (Springer, Heidelberg, 1997).

[9] L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1974).

[10] U. Rössler, *Solid State Theory* (Springer, Heidelberg, 2004).

[11] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).

[12] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).

[13] G. Bastard, Phys. Rev. B **24**, 5693 (1981).

[14] M. Altarelli, U. Ekenberg, and A. Fasolino, Phys. Rev. B **32**, 5138 (1985).

[15] M. G. Burt, J. Phys.: Condens. Matter **4**, 6651 (1992).

[16] M. G. Burt, J. Phys.: Condens. Matter **11**, R53 (1999).

[17] B. A. Foreman, Phys. Rev. B **48**, 4964 (1993).

[18] B. A. Foreman, Phys. Rev. B **54**, 1909 (1996).

[19] B. A. Foreman, Phys. Rev. B **56**, R12748 (1997).

[20] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors* (Springer, Berlin, 2003).

[21] P. Löwdin, J. Chem. Phys. **19**, 1396 (1951).

[22] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**, 5815 (2001).

[23] G. Dresselhaus, A. F. Kip, and C. Kittel, Phys. Rev. **98**, 368 (1955).

[24] *Semiconductors: Intrinsic Poperties of Group IV Elements and III-V, II-VI and I-VII Compounds*, Landolt-Börnstein, New Series, Group III, Vol. 22, Pt. A, edited by O. Madelung (Springer, Berlin, 1987).

[25] P. Lawaetz, Phys. Rev. B **4**, 3460 (1971).

[26] M. F. H. Schuurmans and G. W. 't Hooft, Phys. Rev. B **31**, 8041 (1985).

[27] A. V. Rodina, A. Yu. Alekseev, A. L. Efros, M. Rosen, and B. K. Meyer, Phys. Rev. B **65**, 125302 (2002).

[28] T. Andlauer, Diploma thesis, Technische Universität München (2004).

[29] S. R. White and L. J. Sham, Phys. Rev. Lett. **47**, 879 (1981).

[30] R. E. Bank, J. F. Bürgler, W. Fichtner, and R. K. Smith, Numer. Math. **58**, 185 (1990).

[31] S. L. Cunningham, Phys. Rev. B **10**, 4988 (1974).

[32] J. F. Cooke and R. F. Wood, Phys. Rev. B **5**, 1276 (1972).

[33] G. Wiesenekker, G. te Velde, and E. J. Baerends, J. Phys. C: Solid State Phys. **21**, 4263 (1988).

[34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 90* (Cambridge University Press, Cambridge, U.K., 1996).

[35] A. Trellakis, Ph.D. thesis, University of Illinois at Urbana-Champaign (2000).

[36] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[37] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[38] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[39] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, Oxford, 1959).

[40] J. F. Nye, *Physical Properties of Crystals* (Oxford University Press, 1985).

[41] C. Pryor, J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger, J. Appl. Phys. **83**, 2548 (1998).

[42] S.-H. Park and S. L. Chuang, Phys. Rev. B **59**, 4725 (1999).

[43] M. A. Migliorato, A. G. Cullis, M. Fearn, and J. H. Jefferson, Phys. Rev. B **65**, 115316 (2002).

[44] Q. Xie, A. Madhukar, P. Chen, and N. P. Kobayashi, Phys. Rev. Lett. **75**, 2542 (1995).

[45] C. Y. P. Chao and S. L. Chuang, Phys. Rev. B **46**, 4110 (1992).

[46] D. L. Smith and C. Mailhiot, J. Appl. Phys. **63**, 2717 (1988).

[47] T. Eißfeller, Diploma thesis, Technische Universität München (2008).

[48] See http://www.netlib.org/blas/ for obtaining the BLAS libraries and related publications.

[49] See http://www.netlib.org/lapack/ for obtaining the LAPACK libraries and related publications.

[50] R. S. Varga, *Matrix Iterative Analysis* (Springer, Berlin, 2000).

[51] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods* (SIAM, Philadelphia, 1994).

[52] M. R. Hestenes and E. Stiefel, J. Res. Nat. Bur. Stand. **49**, 409 (1952).

[53] T. F. Chan and T. Szeto, SIAM J. Sci. Comput. **17**, 1491 (1996).

[54] C. Lanczos, J. Res. Nat. Bur. Stand. **49**, 33 (1952).

[55] J. A. Meijernick and H. A. van der Vorst, Math. Comp. **31**, 148 (1977).

[56] T. Dupont, R. P. Kendall, and H. H. Rachford Jr., SIAM J. Numerical Analysis **5**, 559 (1968).

[57] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide* (SIAM, Philadelphia, 2000).

[58] See http://www.caam.rice.edu/software/ARPACK/ for obtaining the ARPACK libraries and related publications.

[59] G. L. G. Sleijpen and H. A. van der Vorst, SIAM J. Matrix Anal. Appl. **17**, 401 (1996).

[60] F. Stern, J. Comput. Phys. **6**, 56 (1970).

[61] A. Trellakis, A. T. Galick, A. Pacelli, and U. Ravaioli, J. Appl. Phys. **81**, 7880 (1997).

[62] S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, and P. Vogl, IEEE Trans. Electron Devices **54**, 2137 (2007).

[63] M. Graf and P. Vogl, Phys. Rev. B **51**, 4940 (1995).

[64] T. B. Boykin, R. C. Bowen, and G. Klimeck, Phys. Rev. B **63**, 245314 (2001).

[65] B. A. Foreman, Phys. Rev. B **66**, 165212 (2002).

[66] A. V. Nenashev, A. V. Dvurechenskii, and A. F. Zinovieva, Phys. Rev. B **67**, 205301 (2003).

[67] W. Sheng and A. Babinski, Phys. Rev. B **75**, 033316 (2007).

[68] P. K. Misra and L. M. Roth, Phys. Rev. **177**, 1089 (1968).

[69] N. C. Das and P. K. Misra, Phys. Rev. B **4**, 225 (1971).

[70] W. Cai and G. Galli, Phys. Rev. Lett. **92**, 186402 (2004).

[71] M. Governale and C. Ungarelli, Phys. Rev. B **58**, 7816 (1998).

[72] C. E. Pryor and M. E. Flatté, Phys. Rev. Lett. **96**, 026804 (2006), Phys. Rev. Lett. **99**, 179901(E) (2007).

[73] R. Peierls, Z. Phys. **80**, 763 (1933).

[74] L. M. Roth, B. Lax, and S. Zwerdling, Phys. Rev. **114**, 90 (1959).

[75] A. A. Kiselev, K. W. Kim, and E. Yablonovitch, Phys. Rev. B **64**, 125303 (2001).

[76] I. Montvay and G. Münster, *Quantum Fields on a Lattice* (Cambridge University Press, Cambridge, England, 1997).

[77] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).

[78] H. W. van Kesteren, E. C. Cosman, W. A. J. A. van der Poel, and C. T. Foxon, Phys. Rev. B **41**, 5283 (1990).

[79] M. J. Snelling, E. Blackwood, C. J. McDonagh, R. T. Harley, and C. T. B. Foxon, Phys. Rev. B **45**, 3922 (1992).

[80] N. J. Traynor, R. T. Harley, and R. J. Warburton, Phys. Rev. B **51**, 7361 (1995).

[81] X. Marie, T. Amand, P. Le Jeune, M. Paillard, P. Renucci, L. E. Golub, V. D. Dymnikov, and E. L. Ivchenko, Phys. Rev. B **60**, 5811 (1999).

[82] M. Oestreich, A. P. Herbele, W. W. Rühle, R. Nötzel, and K. Ploog, Europhys. Lett. **31**, 399 (1995).

[83] M. Bayer, A. Kuther, A. Forchel, A. Gorbunov, V. B. Timofeev, F. Schäfer, J. P. Reithmaier, T. L. Reinecke, and S. N. Walck, Phys. Rev. Lett. **82**, 1748 (1999).

[84] M. Bayer, O. Stern, A. Kuther, and A. Forchel, Phys. Rev. B **61**, 7273 (2000).

[85] T. Nakaoka, T. Saito, J. Tatebayashi, and Y. Arakawa, Phys. Rev. B **70**, 235337 (2004).

[86] M. T. Björk, A. Fuhrer, A. E. Hansen, M. W. Larsson, L. E. Fröberg, and L. Samuelson, Phys. Rev. B **72**, 201307(R) (2005).

[87] T. P. Mayer Alegre, F. G. G. Hernández, A. L. C. Pereira, and G. Medeiros-Ribeiro, Phys. Rev. Lett. **97**, 236402 (2006).

[88] V. Fock, Z. Phys. **47**, 446 (1928), C. G. Darwin, Proc. Camb. Phil. Soc. **27**, 86 (1930).

[89] H.-R. Trebin, U. Rössler, and R. Ranvaud, Phys. Rev. B **20**, 686 (1979).

[90] C. Sliwa and T. Dietl, Phys. Rev. B **74**, 245215 (2006).

[91] A. De and C. E. Pryor, Phys. Rev. B **76**, 155321 (2007).

[92] U. Ekenberg and M. Altarelli, Phys. Rev. B **32**, 3712 (1985).

[93] M. W. Larsson, J. B. Wagner, M. Wallin, P. Hakansson, L. E. Fröberg, L. Samuelson, and L. R. Wallenberg, Nanotech. **18**, 015504 (2007).

[94] G. Hendorfer and J. Schneider, Semicond. Sci. Technol. **6**, 595 (1991).

[95] J. J. Finley, M. Sabathil, P. Vogl, G. Abstreiter, R. Oulton, A. I. Tartakovskii, D. J. Mowbray, M. S. Skolnick, S. L. Liew, A. G. Cullis, et al., Phys. Rev. B **70**, 201308(R) (2004).

[96] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, and A. C. Gossard, Science **309**, 2180 (2005).

[97] M. Kroutvar, Y. Ducommun, D. Heiss, M. Bichler, D. Schuh, G. Abstreiter, and J. J. Finley, Nature (London) **432**, 81 (2004).

[98] F. H. L. Koppens, C. Buizert, K. J. Tielrooij, I. T. Vink, K. C. Nowack, T. Meunier, L. P. Kouwenhoven, and L. M. K. Vandersypen, Nature (London) **442**, 766 (2006).

[99] B. E. Kane, Nature (London) **393**, 133 (1998).

[100] G. Salis, Y. Kato, K. Ensslin, D. C. Driscoll, A. C. Gossard, and D. D. Awschalom, Nature (London) **414**, 619 (2001).

[101] Y. Kato, R. C. Myers, D. C. Driscoll, A. C. Gossard, J. Levy, and D. D. Awschalom, Science **299**, 1201 (2003).

[102] J. H. H. Rietjens, G. W. W. Quax, C. A. C. Bosco, R. Nötzel, A. Yu. Silov, and B. Koopmans, J. Appl. Phys. **103**, 07B116 (2008).

[103] M. F. Doty, M. Scheibner, I. V. Ponomarev, E. A. Stinaff, A. S. Bracker, V. L. Korenev, T. L. Reinecke, and D. Gammon, Phys. Rev. Lett. **97**, 197202 (2006).

[104] C. F. Destefani, S. E. Ulloa, and G. E. Marques, Phys. Rev. B **70**, 205315 (2004).

[105] A. S. Bracker, M. Scheibner, M. F. Doty, E. A. Stinaff, I. V. Ponomarev, J. C. Kim, L. J. Whitman, T. L. Reinecke, and D. Gammon, Appl. Phys. Lett. **89**, 233110 (2006).

[106] H. J. Krenner, M. Sabathil, E. C. Clark, A. Kress, D. Schuh, M. Bichler, G. Abstreiter, and J. J. Finley, Phys. Rev. Lett. **94**, 057402 (2005).

[107] P. Offermans, P. M. Koenraad, J. H. Wolter, K. Pierz, M. Roy, and P. A. Maksym, Phys. Rev. B **72**, 165332 (2005).

[108] M. F. Doty, J. I. Climente, M. Korkusinski, M. Schreibner, A. S. Bracker, P. Hawrylak, and D. Gammon, arXiv:0804.3097 (unpublished).

[109] M. Grundmann, O. Stier, and D. Bimberg, Phys. Rev. B **52**, 11969 (1995).

[110] J. Pingenot, C. E. Pryor, and M. E. Flatté, Appl. Phys. Lett. **92**, 222502 (2008).

[111] J. B. Boos, B. R. Bennett, N. A. Papanicolaou, M. G. Ancona, J. G. Champlain, R. Bass, and B. V. Shanabrook, Electron. Lett. **43**, 834 (2007).

[112] R. Ascazubi, I. Wilke, K. J. Kim, and P. Dutta, Phys. Rev. B **74**, 075323 (2006).

[113] W. Xu, X. F. Wei, and J. Zhang, Appl. Phys. Lett. **92**, 162108 (2008).

[114] A. J. L. Poulter, M. Lakrimi, R. J. Nicholas, N. J. Mason, and P. J. Walker, Phys. Rev. B **59**, 10785 (1999).

[115] H. Mohseni, E. Michel, J. Sandoen, M. Razeghi, W. Mitchel, and G. Brown, Appl. Phys. Lett. **71**, 1403 (1997).

[116] G. Liu and S.-L. Chuang, Phys. Rev. B **65**, 165220 (2002).

[117] Y. Vasilyev, S. Suchalkin, K. von Klitzing, B. Meltser, S. Ivanov, and P. Kop'ev, Phys. Rev. B **60**, 10636 (1999).

[118] M. Altarelli, Phys. Rev. B **28**, 842 (1983).

[119] W. Xu, P. A. Folkes, and G. Gumps, J. Appl. Phys. **102**, 033703 (2007).

[120] I. Lapushkin, A. Zakharova, S. T. Yen, and K. A. Chao, J. Phys.: Condens. Matter **16**, 4677 (2004).

[121] I. Semenikhin, A. Zakharova, and K. A. Chao, Phys. Rev. B **77**, 113307 (2008).

[122] L. L. Chang, G. A. Sai-Halasz, L. Esaki, and R. L. Aggarwal, J. Vac. Sci. Technol. **19**, 589 (1981).

[123] J. A. Majewski and P. Vogl, Phys. Rev. B **35**, 9666 (1987).

[124] A. Baldereschi, S. Baroni, and R. Resta, Phys. Rev. Lett. **61**, 734 (1988).

[125] C. G. Van de Walle and R. M. Martin, Phys. Rev. B **35**, 8154 (1987).

[126] S.-H. Wei and A. Zunger, Appl. Phys. Lett. **72**, 2011 (1998).

[127] G. Feher, J. C. Hensel, and E. A. Gere, Phys. Rev. Lett. **5**, 309 (1960).

[128] H. Neubrand, Phys. Status Solidi (b) **90**, 301 (1978).

[129] A. R. Stegner, T. Andlauer, H. Tezuka, M. Stutzmann, K. M. Itoh, and M. S. Brandt, in preparation.

[130] D. Karaiskaj, G. Kirczenow, M. L. W. Thewalt, R. Buczko, and M. Cardona, Phys. Rev. Lett. **90**, 016404 (2003).

[131] J. Knipping, H. Wiggers, B. Rellinghaus, P. Roth, D. Konjhodzic, and C. Meier, J. Nanosci. Nanotechnol. **4**, 1039 (2004).

[132] S. C. Erwin, L. Zu, M. I. Haftel, A. L. Efros, T. A. Kennedy, and D. J. Norris, Nature (London) **436**, 91 (2005).

[133] M. Fujii, A. Mimura, S. Hayashi, Y. Yamamoto, and K. Murakami, Phys. Rev. Lett. **89**, 206805 (2002).

[134] L. W. Wang and A. Zunger, Phys. Rev. Lett. **73**, 1039 (1994).

[135] R. N. Pereira, A. R. Stegner, T. Andlauer, K. Klein, M. S. Brandt, M. Stutzmann, and H. Wiggers, submitted to Phys. Rev. Lett.

[136] S. T. Pantelides and C. T. Sah, Phys. Rev. B **10**, 621 (1974).

[137] V. A. Belyakov and V. A. Burdov, J. Phys.: Condens. Matter **20**, 025213 (2008).

[138] S. Öğüt, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. Lett. **79**, 1770 (1997).

[139] V. A. Belyakov and V. A. Burdov, Phys. Rev. B **76**, 045335 (2007).

[140] T. L. Chan, M. L. Tiago, E. Kaxiras, and J. R. Chelikowsky, Nano Lett. **8**, 596 (2008).

[141] G. Feher, Phys. Rev. **114**, 1219 (1959).

[142] S. L. Chuang and C. S. Chang, Phys. Rev. B **54**, 2491 (1996).

[143] I. Vurgaftman and J. R. Meyer, J. Appl. Phys. **94**, 3675 (2003).

[144] F. Mireles and S. E. Ulloa, Phys. Rev. B **60**, 13659 (1999).

[145] J. A. Majewski, private communication.

[146] K. Yang, T. Anan, and L. J. Schowalter, Appl. Phys. Lett. **65**, 2789 (1994).

Danksagung

An dieser Stelle möchte ich allen danken, die mich während dieser Arbeit unterstützt haben. Zunächst gilt mein besonderer Dank Prof. Dr. Peter Vogl für die Ermöglichung dieser Dissertation, seine fortwährende Betreuung und großzügige Förderung, die gewährte Freiheit bei der Interpretation des Themas, sowie die Bereitstellung einer idealen Arbeitsumgebung.

Weiterhin danke ich Dr. Tobias Zibold für die intensive Zusammenarbeit bei der Entwicklung von nextnano++, die zahlreichen Diskussionen über physikalische Problemstellungen und das sehr gute Arbeitsklima in unserem gemeinsamen Büro. Ebenso danke ich Dr. Alex Trellakis für die gemeinsame Arbeit an nextnano++, aus der ich viel über numerische Methoden gelernt habe.

Meinen weiteren Kollegen von T33, Stefan Birner, Tillmann Kubis, Christoph Schindler, Thomas Eißfeller, Peter Greck und Dr. Reinhard Scholz danke ich für die äußerst angenehme Arbeitsatmosphäre und die zahlreichen konstruktiven Diskussionen. Besonders erwähnen möchte ich hierbei noch Stefans unermüdliches Engagement für nextnano sowie Peters und Thomas stets kompetente Hilfe bei Softwarefragen.

Mein Dank gilt auch unseren Sekretärinnen Veronika Enter und Liane Lindner für ihre große Hilfsbereitschaft bei allen verwaltungs und organisatorischen Problemen.

Desweiteren danke ich André Stegner für die hervorragende und äußerst produktive Zusammenarbeit zum Thema Störstellen in Silizium, die stets in einer sehr offenen und angenehmen Atmosphäre stattfand.

Für die gute Zusammenarbeit mit den weiteren experimentellen Gruppen, sowie bei allen anderen Kollegen und Mitarbeitern am Walter Schottky Institut möchte ich mich recht herzlich für die interessanten und schönen Jahre bedanken, die ich an diesem Institut verbringen durfte.

Vor allem aber gilt mein Dank meinen Eltern, die mir durch ihre Unterstützung erst diese Ausbildung bis hin zur Promotion ermöglichten, sowie meiner Freundin Barbara für ihre geduldige und liebevolle Unterstützung.