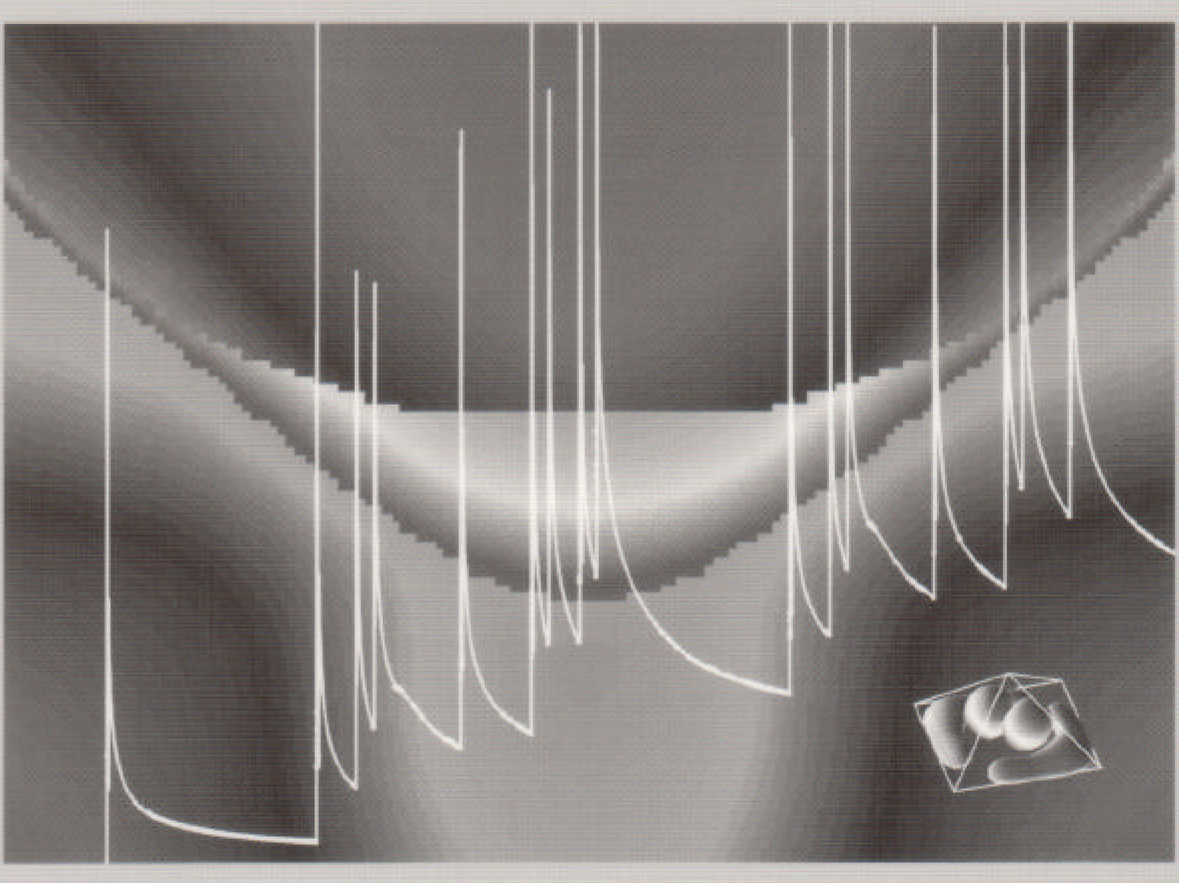
Berlin Studies in Solid State Physics 7

Oliver Stier

Electronic and Optical Properties of Quantum Dots and Wires



W&T
WISSENSCHAFT &
TECHNIK VERLAG

Berlin Studies in Solid State Physics 7

Theoretical predictions of the electronic properties of semiconductor quantum dots and wires provide a link between investigations of the structural and optical properties of such systems. A complex numerical model enables the simulation of the impact which shape, chemical composition, strain, and piezoelectricity have on the electronic and optical properties of nanostructures, eventually leading to a comprehensive understanding of a manifold of prevailing experimental data. The eight-band **k·p** model provides a fast and transparent connection between the electronic structure of quantum dots or wires and the bulk properties of the constituent materials. Absorption spectra and fine-structure properties of few-particle states are subsequently calculated within a coherent framework. The predictive power of the method is demonstrated by comparison with previous second-principles calculations and various experiments, yielding outstanding agreement between calculated and measured quantities. As specific nanostructures, capped InAs quantum dots on GaAs (001) grown in the Stranski-Krastanow mode, and V-groove quantum wires of InGaAs fabricated by growth on patterned substrates of AlGaAs or InP are modeled.

Editors: C. Thomsen, D. Bimberg, M. Dähne, W. Richter
Institut für Festkörperphysik, Technische Universität Berlin

ISBN 3-89685-366-X
ISSN 1615-9284

This thesis appears as printed book at the publishing company

Wissenschaft & Technik Verlag
Dresdener Strasse 26
D-10999 Berlin
Germany

and is, as well as any part of it, protected by copyright.

All rights, including translation, reprinting, reproduction, and distribution are with the publishing company. No part of this work may be reproduced in whatsoever manner, electronically processed, or distributed without written permission of the publishing company.

This file appears by courtesy of the publishing company. It is identical with the print edition, and must not be changed.

Contact the author: oliver.stier@alumni.tu-berlin.de

Electronic and Optical Properties of Quantum Dots and Wires

vorgelegt von
Dipl.-Phys. Oliver Stier
aus Berlin

Vom Fachbereich 4 – Physik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
– Dr. rer. nat. –
genehmigte Dissertation

Promotionsausschuß:

Vorsitzender:	Prof. Dr. Hans-Joachim Eichler
Berichter:	Prof. Dr. Dieter Bimberg
	Prof. Dr. Ortwin Hess
	Priv.-Doz. Dr. Axel Hoffmann

Tag der wissenschaftlichen Aussprache: 20.10.2000

Berlin 2000

D 83

Zusammenfassung

Die elektronischen und optischen Eigenschaften von verspannten Halbleiter-Quantenfäden und -Quantenpunkten realistischer Form und chemischer Zusammensetzung werden theoretisch untersucht, d. h. im Rahmen des auf Verspannungen erweiterten 8-Band-**k·p**-Modells unter Einbeziehung des piezoelektrischen Effekts berechnet. Ausgehend von der elektronischen Struktur werden Absorptionsspektren sowie die Eigenschaften von Mehrteilchenzuständen in solchen Nanostrukturen berechnet und, soweit möglich, mit experimentellen Ergebnissen verglichen. Bei derartigen Vergleichen ergibt sich in allen Fällen eine sehr gute Übereinstimmung zwischen berechneten und gemessenen Eigenschaftsgrößen.

Hauptergebnis der Arbeit ist die Etablierung eines theoretischen Konzepts zur realitätsnahen Modellierung von Quantenfäden und -punkten, dessen Genauigkeit anhand eines Vergleichs mit der jüngsten Literatur entnommenen, empirischen Pseudopotentialberechnungen quantifiziert wurde. Da das Modell keine zu variierenden Parameter enthält, können optische Eigenschaften gegebener Strukturen tatsächlich vorhergesagt und nicht etwa nur experimentellen Befunden angepaßt werden.

Anhand zweier Beispiele, nämlich (a) im Stranski-Krastanow-Modus gewachsener InAs/GaAs-Quantenpunkte und (b) V-Graben-Quantenfäden im InGaAs/-AlGaAs-Materialsystem, wird die nichttriviale Korrelation zwischen den strukturellen und optischen Eigenschaften niederdimensionaler Halbleiterstrukturen aufgezeigt. Der dominierende Einfluß der inhomogenen Verspannung und der piezoelektrischen Felder auf die Struktur und Polarisationseigenschaften von Absorptionsspektren geht aus der detailgetreuen Berechnung anschaulich hervor. Vereinfachte Modelle, welche die Geometrie im Realraum durch elementare Formen wie z. B. Kugeln oder Quader annähern, oder die Bandstruktur der betreffenden Halbleiter in entkoppelten Modellen wie der Effektivmasse-Näherung behandeln, sind in der Vergangenheit zwar häufigst benutzt worden, erweisen sich hier jedoch als völlig ungeeignet.

Die besondere Detailtreue des Modells gestattet auch die Untersuchung von Mehrteilchen-Zuständen in Quantenpunkten, wie Exzitonen, geladenen Exzitonen (Trionen) und Biexzitonen. Mittels der Konfigurations-Wechselwirkungs-Methode kann die empfindliche Balance zwischen der Coulomb-Wechselwirkung und dem Austausch-Korrelations-Effekt genau berechnet werden. Es zeigt sich, daß die Bindungsenergien solcher Zustände je nach Gestalt, Zusammensetzung und Größe der Quantenpunkte positiv oder negativ sein können. Ergebnisse von Einzel-Quantenpunkt-Spektroskopie-Experimenten bestätigen die Existenz dieser beiden Möglichkeiten. Auch die für Exzitonen berechnete Feinstruktur-Aufspaltung infolge von Austausch-Effekten stimmt quantitativ mit an vergleichbaren Quantenpunktstrukturen gemessenen Linienabständen überein.

Insgesamt zeigt sich, daß die eingesetzte Modellierungsmethode quantitativ stichhaltige Vorraussagen für konkrete Strukturen liefert und, ergänzend zu experimentellen Methoden, zur Aufklärung der physikalischen Zusammenhänge zwischen strukturellen und elektronischen Eigenschaften eingesetzt werden kann. Die für systematische Untersuchungen erforderliche Rechenkapazität wird durch die Implementierung für Parallelrechner vom Typ Cray T3E gewährleistet.

Parts of this work have been published in

- O. Stier, M. Grundmann, D. Bimberg,
  Electronic and optical properties of strained quantum dots modeled by eight-band $\mathbf{k}\cdot\mathbf{p}$ *theory*,
  Phys. Rev. B **59**, 5688 (1999).

- R. Heitz, O. Stier, I. Mukhametzanov, A. Madhukar, D. Bimberg,
  Quantum size effect in self-organized InAs/GaAs quantum dots,
  Phys. Rev. B **62**, 11017 (2000).

- R. Heitz, I. Mukhametzhanov, O. Stier, A. Madhukar, D. Bimberg,
  Enhanced Polar Exciton-LO-Phonon Interaction in Quantum Dots,
  Phys. Rev. Lett. **83**, 4654 (1999).

- M. Grundmann, O. Stier, S. Bognár, C. Ribbat, F. Heinrichsdorff, D. Bimberg,
  Optical Properties of Self-Organized Quantum Dots: Modeling and Experiments,
  phys. stat. sol. (a) **178**, 255 (2000).

- R. Heitz, I. Mukhametzhanov, O. Stier, A. Madhukar, D. Bimberg,
  Phonon-assisted polar exciton transitions in self-organized InAs/GaAs quantum dots,
  Physica E **7**, 398 (2000).

- C. M. A. Kapteyn, F. Heinrichsdorff, O. Stier, R. Heitz, M. Grundmann, N. D. Zakharov, D. Bimberg, P. Werner,
  Electron escape from InAs quantum dots,
  Phys. Rev. B **60**, 14265 (1999).

- O. Stier, M. Grundmann, D. Bimberg,
  Inter- and intraband transitions in strained quantum dots modeled in eight-band $\mathbf{k}\cdot\mathbf{p}$ *theory*,
  Proc. 24th International Conference on the Physics of Semiconductors, Jerusalem, Israel, 1998, ed. by D. Gershoni (World Scientific, Singapore, 1999), No. 884.

- C. M. A. Kapteyn, F. Heinrichsdorff, O. Stier, M. Grundmann, D. Bimberg,
  Electron Emission from InAs quantum dots,
  Proc. 24th International Conference on the Physics of Semiconductors, Jerusalem, Israel, 1998, ed. by D. Gershoni (World Scientific, Singapore, 1999), No. 1339.

- O. Stier, D. Bimberg,
  Modeling of strained quantum wires using eight-band $\mathbf{k}\cdot\mathbf{p}$ *theory*,
  Phys. Rev. B **55**, 7726 (1997).

- O. Stier, M. Grundmann, D. Bimberg,
  Eight-band $\mathbf{k}\cdot\mathbf{p}$ *analysis of pseudomorphic quantum wires: The InGaAs/AlGaAs case*,
  Proc. 23rd International Conference on the Physics of Semiconductors, Berlin, Germany, 1996, ed. by M. Scheffler and R. Zimmermann (World Scientific, Singapore, 1996), p. 1177.

- O. Stier, V. Türck, M. Kappelt, D. Bimberg,
  First observation of symmetry breaking in strained $In_{0.7}Ga_{0.3}As/InP$ *V-groove quantum wires*,
  Physica E **2**, 969 (1998).

- M. Grundmann, O. Stier, A. Schliwa, D. Bimberg,
  Electronic structure of cleaved-edge-overgrowth strain-induced quantum wires,
  Phys. Rev. B **61**, 1744 (2000).

- E. Martinet, M.-A. Dupertuis, F. Reinhardt, G. Biasiol, E. Kapon, O. Stier, M. Grundmann, D. Bimberg,
  Separation of strain and quantum-confinement effects in the optical spectra of quantum wires,
  Phys. Rev. B **61**, 4488 (2000).

- E. L. Martinet, M.-A. Dupertuis, E. Kapon, O. Stier, M. Grundmann, D. Bimberg,
  Separation of strain and confinement effects in the photoluminescence excitation spectra of InGaAs/AlGaAs V-groove quantum wires,
  Proc. 24th International Conference on the Physics of Semiconductors, Jerusalem, Israel, 1998, ed. by D. Gershoni (World Scientific, Singapore, 1999), No. 1449.

- M. Kappelt, V. Türck, O. Stier, D. Bimberg, D. Stenkamp,
  Quantization effects of InGaAs/InP-quantum wires grown on patterned substrates,
  Proc. 9th International Conference on Indium Phosphide and Related Materials, Hyannis, USA, IEEE, p. 83 (1997).

Contents

I Electronic Structure Calculation by Eight-Band k·p Theory 1

1 Introduction 3

2 The Eight-Band k·p Model Including Strain 5

2.1 The eight-band **k·p** Hamiltonian 8

2.2 Heterostructures 11

2.3 Material parameters 11

3 Quantum-Confined Heterostructures 17

3.1 The model quantum dot 18

3.1.1 Strain effects 18

3.1.2 Piezoelectricity 20

3.1.3 Bandstructure 21

3.2 Predicted quantum dot properties 24

3.2.1 Neglection of the piezoelectric effect 24

3.2.2 Inclusion of piezoelectricity 29

3.2.3 Smaller quantum dots 29

3.3 Resumé 30

II Electronic Structure of Quantum Dots and Wires 33

4 Electronic Structure of Quantum Dots 35

4.1 Calculation of strain 39

4.1.1 Continuum mechanical model 39
4.1.2 Valence force field model 39
4.2 Piezoelectricity 45
4.3 Bound single-particle states 47
4.3.1 Results for strain according to the CM model 47
4.3.2 Impact of input values 53
4.3.3 Results for strain according to the VFF model 55
4.4 Dipole transitions 56
4.4.1 Time reversal symmetry 59

5 Few-Particle States in Quantum Dots 61

5.1 LDA in the Hartree approximation 63
5.2 Configuration interaction calculations 66
5.2.1 Fermion character and spin 67
5.2.2 Dielectric screening 68
5.2.3 Computational expense 69
5.2.4 Exciton in the Hartree approximation 70
5.2.5 Excitonic finestructure 72
5.2.6 Trion ground states 80
5.2.7 Biexciton energy levels 81

6 Optical Properties of Quantum Dots 85

6.1 Calculated spectra 85
6.1.1 Excitonic transitions 86
6.1.2 Intraband transitions 88
6.2 Exciton phonon interaction 91
6.3 Comparison with experiments 93
6.4 Resumé 101

7 Bandstructure in Quantum Wires 103

7.1 Eight-band $\mathbf{k}\cdot\mathbf{p}$ model of strained quantum wires 105
7.2 Specific quantum wire structures 112

7.2.1 $In_{0.15}Ga_{0.85}As/Al_{0.15}Ga_{0.85}As$ V-groove quantum wires . . . 113
7.2.2 $In_{0.7}Ga_{0.3}As/InP$ V-groove quantum wires 113
7.2.3 Strained cleaved-edge overgrowth quantum wires 114

III Numerical Procedures 115

8 Finite Differences Method 117

9 Calculation of Confined States 119
9.1 Hamiltonian fill-in structures 121
9.2 Efficiency of prevailing eigensolvers 123
9.2.1 Ordering scheme for iterative eigensolvers 124
9.2.2 Power method 126
9.2.3 Spectral transformation 127
9.2.4 Simultaneous iteration 129
9.2.5 Krylov methods 130
9.2.6 Three-term recurrences 131
9.2.7 Convergence rate of the power method 133
9.2.8 Chebyshev iteration 133
9.2.9 Second order molecular dynamics 135
9.2.10 Modified second order molecular dynamics 137
9.2.11 Modified first order molecular dynamics 141
9.2.12 Comparison of methods 142
9.2.13 Preconditioning 143
9.2.14 Generalized Davidson algorithm 146
9.2.15 Residual minimization 151
9.3 Improved algorithm for interior eigenpairs 152
9.3.1 Benchmarks 159
9.3.2 Programming considerations 162
9.4 Resumé 166
9.5 Summary and Outlook 167

Part I

Electronic Structure Calculation by Eight-Band k·p Theory

Chapter 1

Introduction

The continuous progress in epitaxial growth and patterning technology offers today unprecedented possibilities to fabricate isolated semiconductor heterostructures which exhibit quantum confinement of charge carriers in two and three spatial ($\mathbf{r}$) dimensions. The $\mathbf{k}$ vector spaces of confined (bound) electrons and holes are zero-dimensional (0D) for quantum dots (QDs) [1–5] and one-dimensional (1D) for quantum wires (QWRs) [6–8].

Theoretical predictions of the electronic properties of semiconductor QDs and QWRs, as presented in this work, are important since they provide a link between investigations of the structural and optical properties of such systems. Complex numerical models are prerequisite for the simulation of the non-trivial impact of the real structure (e. g., shape, chemical composition, inhomogeneous strain, piezoelectricity) on the electronic and optical properties, eventually leading to a comprehensive understanding of the manifold of experimental data available, see, for instance, Ref. [9–12] for QDs and Ref. [13–16] for QWRs. Quantitatively correct predictions will also ultimately provide guidance to tailor the properties of optoelectronic devices based on self-organized QDs [17, 18] or QWRs [8].

The realization of this potential depends on the predictive power of the chosen theoretical model. It is commonly agreed that the inhomogeneous strain inherently connected to the formation of strained QDs [19], e. g., within the InAs/GaAs material systems, strongly influences their electronic structure. Band mixing effects, being caused by confinement in general and enhanced by the high dimension and low symmetry of the confinement and by strain effects, are of equal significance. These insights promoted the use of various perturbation [20–23] and multiband models [24–32] combined with adiabatic treatments of the strain effects [33,34], as well as tight-binding [35] and pseudopotential calculations [36–39], whose fully numerical implementations allow to treat realistic geometrical and chemical properties of the dots. Comparably detailed bandstructure calculations of strained QWRs, taking into account strain, band-mixing, and piezoelectricity have for the first time been reported [40] in the frame of this work.

Among the above mentioned calculation methods the eight-band $\mathbf{k}\cdot\mathbf{p}$ model provides, at reasonable computational cost, a fast and transparent connection between the electronic structure of QDs and certain bulk properties of the constituent materials, easing to take into account arbitrary QD shapes and material compositions, as well as strain, piezoelectricity, and band mixing. This renders the eight-band $\mathbf{k}{\cdot}\mathbf{p}$ model attractive for a realistic analysis of large QD structures.

In this work, eight-band $\mathbf{k}\cdot\mathbf{p}$ theory including strain and piezoelectricity is applied to the calculation of the electronic and optical properties of strained quantum dots and quantum wires of arbitrary shape and chemical composition. The electronic structure, absorption spectra, and fine-structure properties of few-particle states are calculated within one coherent framework. As it is mandatory to assume realistic bulk properties of the constituent semiconductor materials, the bandstructures near the Brillouin zone center are described by the eight-band $\mathbf{k}\cdot\mathbf{p}$ Kane model including the conduction band, as well as the heavy hole, light hole, and spin-orbit split off valence bands.

The specific heterostructures treated in this work are self-organized InAs QDs in GaAs, grown by the Stranski-Krastanow mode, and V-groove QWRs of InGaAs fabricated by growth on prepatterned substrates of AlGaAs or InP. The electronic and optical properties of these lattice-mismatched heterostructures are strongly influenced by pseudomorphic strain and piezoelectricity. The effects of inhomogeneous strain on the bandstructure of bulk semiconductors is included via linear deformation potential theory as proposed by Pikus and Bir. The piezoelectric effect is modeled including the image charge effects arising at heterointerfaces. To calculate the strain, two different models are considered: The continuum mechanical approach for cubic materials and the atomistic valence force field model.

The overall intention of this work is to establish and validate an all-numerical treatment allowing the flexible modeling of a variety of low-dimensional nanostructures. The predictive power of the used method is demonstrated by comparison with prevailing second-principles calculations. Comparisons to various experiments demonstrate outstanding agreement between calculated and measured quantities.

Chapter 2

The Eight-Band k·p Model Including Strain

A large part of the fundamental understanding of the electronic structure of semiconductors is due to the empirical pseudopotential (EP) method [41–44]. The form factors (shape parameters) of the pseudopotentials representing the atoms in a semiconductor real lattice are empirically fitted in order to yield selected, either experimentally known properties of the respective semiconductor bulk material, or properties predicted by, e. g., first-principles calculations. As an example, Fig. 2.1 shows the bandstructure of bulk InAs [45] at low temperature as predicted by the EP method [46] together with the resulting, calculated valence band (VB) density of states (DOS) and the VB DOS as derived from X-ray photospectroscopy measurements [47].

For the present work, being dedicated to the electronic properties of low-dimensional structures formed from direct semiconductors with zincblende lattice structure (see Fig. 2.2), mainly that part of the bulk bandstructure is important which is displayed on the left-hand side of Fig. 2.1, i. e. the bandstructure in a region centered at the Γ-point ($\mathbf{k} = \mathbf{0}$) shown in Fig. 2.2. As can be seen from Fig. 2.3, this actually is a rather small fraction of the entire bulk bandstructure, with respect to both the **k**-space extension and the energy range. In case of a thermal population of the bands at non-zero temperature the band edges at the Γ-point play, nevertheless, the most important role for optical transitions in a direct semiconductor. Given a large energy difference between the fundamental transition at Γ and transitions in other valleys (like L or X), the Γ-point bandstructure dominates even at non-thermal populations, e. g. at lasing.

The EP method is able to predict the entire complexity of a bandstructure, like that shown in Fig. 2.3 for GaAs [42], from a relatively small[1] set of form factors.

---

[1] In the order of 10 to 20 parameters are employed to describe strained bulk material [36].

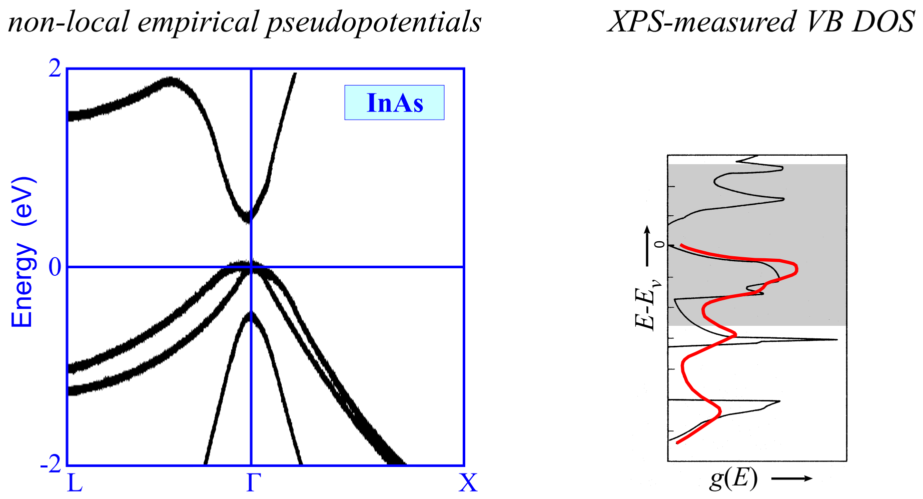


Figure 2.1:

Bulk bandstructure of InAs at 0 K [45] as predicted by the EP method (left), EP calculated VB DOS (right: thin black line), and measured VB DOS from X-ray photospectroscopy (right: thick gray line). The shaded frame on the right-hand side indicates the energy range covered by the bandstructure diagram on the left-hand side.

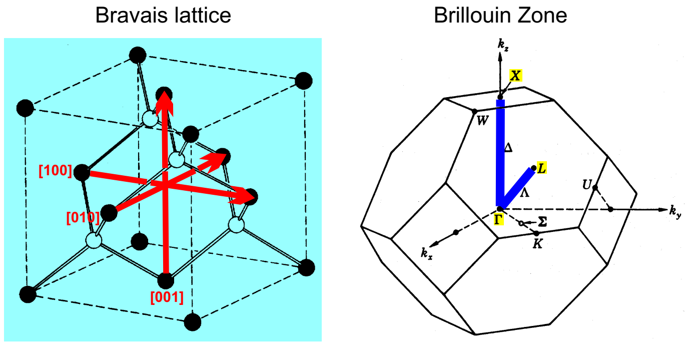


Figure 2.2:

Unit cells of the Bravais lattice (left) and the reciprocal lattice (right, Brillouin zone) of a zincblende-type semiconductor [42]. Black spheres denote the cations and hollow spheres the anions in polar crystals. The thick gray arrows (left) show the principal crystal axes. The thick lines (right) indicate the directions the bandstructure diagram in Fig. 2.1 refers to.

This suggests that the bandstructure part near Γ, being of interest here, may well be described by perturbation theory around $\mathbf{k} = \mathbf{0}$, using a comparable number of (perturbationally defined) parameters. As motivated by Fig. 2.3, an expansion of the Γ-point electron wave function into s- and p-orbital functions appears appropriate. At inclusion of the spin-orbit (SO) interaction, which is responsible for the splitting Δ_0 between the Γ_8 and Γ_7 VBs, a perturbation model requires a basis of eight so called Bloch functions, $s\uparrow$, $x\uparrow$, $y\uparrow$, $z\uparrow$, $s\downarrow$, $x\downarrow$, $y\downarrow$, and $z\downarrow$, where x, y, and z are the p-type Bloch functions refering to the three principal directions in the crystal and the arrows denote the spin. The Hamiltonian with respect to such a basis for the description of electrons in the Γ_6 conduction band (CB) or the Γ_7 or Γ_8 VBs is derived by the so called eight-band $\mathbf{k}\cdot\mathbf{p}$ method [33, 48–53].

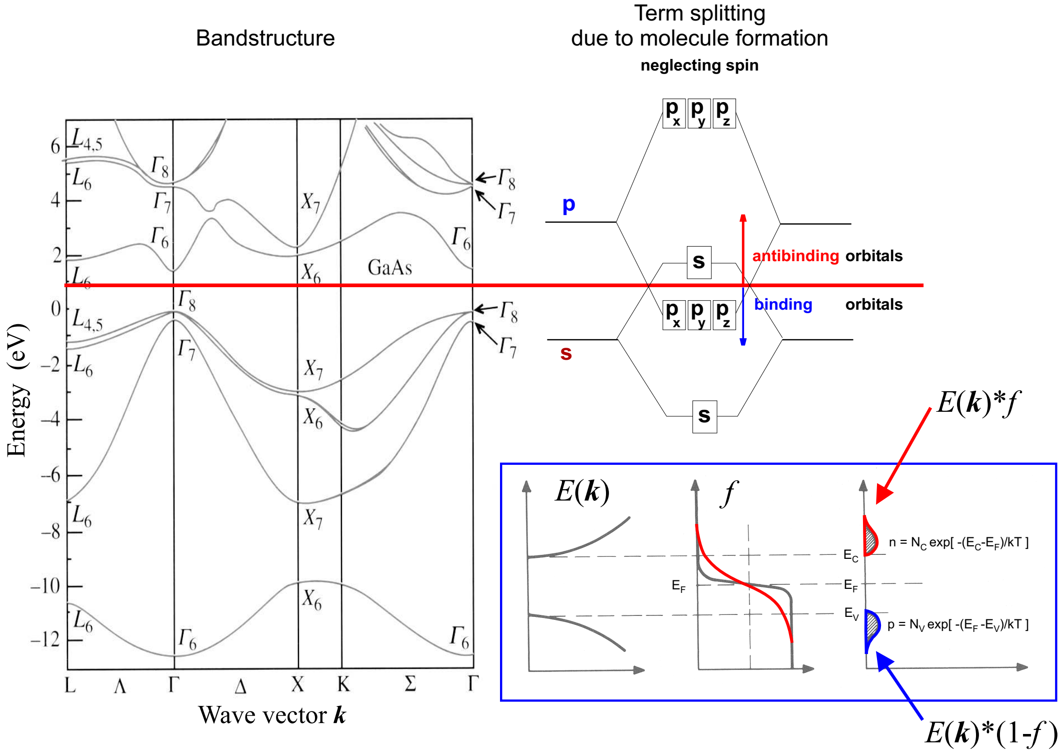

Bandstructure
Term splitting
due to molecule formation
neglecting spin
Energy (eV)
Wave vector k
GaAs
L Λ Γ Δ X K Σ Γ
6
4
2
0
-2
-4
-6
-8
-10
-12
p
s
p_x p_y p_z
s
antibinding orbitals
binding orbitals
E(k)*f
E(k)*(1-f)
E(k)
f
E_F
E_C
E_V
n = N_C exp[-(E_C-E_F)/kT]
p = N_V exp[-(E_F -E_V)/kT]

Figure 2.3:

Left: Bandstructure of GaAs [42] as calculated by the EP method. Right: Qualitative picture of the coarse structure of the left diagram, based on the term splitting occurring during the formation of molecules. In the given case, the splittings of the s and p levels arrange a situation where the upper s state appears at higher energy than the lower p states. The latter dominate the electron wave-function shape in the bulk VB while the former determine the symmetry of the bulk CB. The framed diagram (bottom) sketches the thermal population of the resulting Γ-point bands, showing the CB and VB dispersions $E(\mathbf{k})$, the Fermi distribution f for two different temperatures (light and dark gray curves) and the resulting electron and hole population densities.

2.1 The eight-band k·p Hamiltonian

The key feature of the **k·p** method [48, 54, 55] is the so called envelope-function ansatz based on Bloch's theorem, according to which the electron wave function in a coherent crystal lattice with translation symmetries can be separated into (i) a Bloch part oscillating with the atomic distance as period, and (ii) a smooth envelope function which, at most, exhibits variation on a mesoscopic scale. By Löwdin perturbation theory the rapidly oscillating Bloch functions can be eliminated from the electron Hamiltonian, leaving quantum mechanical equations of motion for the envelope functions only.

In the Kane model [48] the envelope functions ψ belonging to the eight above Bloch functions are ordered as follows:

$$\{\psi_{s\Uparrow}, \psi_{x\Uparrow}, \psi_{y\Uparrow}, \psi_{z\Uparrow}, \psi_{s\Downarrow}, \psi_{x\Downarrow}, \psi_{y\Downarrow}, \psi_{z\Downarrow}\} \quad .$$

With respect to this basis, the eight-band **k·p** Hamiltonian has a 2 × 2 block matrix form visualizing the structure of the spin projections in the presence of SO interaction:

$$\hat{H} = \begin{pmatrix} G(\mathbf{k}) & \Gamma \\ -\overline{\Gamma} & \overline{G}(-\mathbf{k}) \end{pmatrix} \tag{2.1}$$

where G and Γ are 4×4 matrices acting on either the $\Uparrow$ or the $\Downarrow$ envelope functions $(\psi_s, \psi_x, \psi_y, \psi_z)$, and the overlining denotes the complex conjugation. The matrix

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta_0/3 \\ 0 & 0 & 0 & -\mathrm{i}\,\Delta_0/3 \\ 0 & -\Delta_0/3 & \mathrm{i}\,\Delta_0/3 & 0 \end{pmatrix}$$

couples the spin projections $\Uparrow$ and $\Downarrow$ due to the SO interaction causing the splitting between the Γ_8 and Γ_7 VBs by Δ_0 . The matrix G is composed from a potential energy part G_1 , a kinetic energy part G_2 , a SO interaction part G_{SO} , and a strain dependent [33, 49, 52, 56] part G_{st} :

$$G = G_1 + G_2 + G_{\mathrm{SO}} + G_{\mathrm{st}}$$

with

$$G_1 = \begin{pmatrix} E_c & \mathrm{i}\,Pk_x & \mathrm{i}\,Pk_y & \mathrm{i}\,Pk_z \\ -\mathrm{i}\,Pk_x & E_v' & 0 & 0 \\ -\mathrm{i}\,Pk_y & 0 & E_v' & 0 \\ -\mathrm{i}\,Pk_z & 0 & 0 & E_v' \end{pmatrix} ,$$

$$G_2 = \begin{pmatrix} A'\mathbf{k}^2 & Bk_yk_z & Bk_xk_z & Bk_xk_y \\ Bk_yk_z & L'k_x^2 + M(k_y^2+k_z^2) & N'k_xk_y & N'k_xk_z \\ Bk_zk_x & N'k_xk_y & L'k_y^2 + M(k_x^2+k_z^2) & N'k_yk_z \\ Bk_xk_y & N'k_xk_z & N'k_yk_z & L'k_z^2 + M(k_x^2+k_y^2) \end{pmatrix},$$

$$G_{\mathrm{SO}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i}\,\Delta_0/3 & 0 \\ 0 & \mathrm{i}\,\Delta_0/3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$G_{\mathrm{st}} = \begin{pmatrix} a_c(\epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz}) & b'\epsilon_{yz} - \mathrm{i}\,P\epsilon_{xj}k^j & b'\epsilon_{zx} - \mathrm{i}\,P\epsilon_{yj}k^j & b'\epsilon_{xy} - \mathrm{i}\,P\epsilon_{zj}k^j \\ b'\epsilon_{yz} + \mathrm{i}\,P\epsilon_{xj}k^j & \left\{\begin{matrix} l\epsilon_{xx}+ \\ m(\epsilon_{yy}+\epsilon_{zz}) \end{matrix}\right\} & n\epsilon_{xy} & n\epsilon_{xz} \\ b'\epsilon_{zx} + \mathrm{i}\,P\epsilon_{yj}k^j & n\epsilon_{xy} & \left\{\begin{matrix} l\epsilon_{yy}+ \\ m(\epsilon_{xx}+\epsilon_{zz}) \end{matrix}\right\} & n\epsilon_{yz} \\ b'\epsilon_{xy} + \mathrm{i}\,P\epsilon_{zj}k^j & n\epsilon_{xz} & n\epsilon_{yz} & \left\{\begin{matrix} l\epsilon_{zz}+ \\ m(\epsilon_{xx}+\epsilon_{yy}) \end{matrix}\right\} \end{pmatrix} \tag{2.2}$$

In the Einstein sum in Eq. (2.2) $j = x, y, z$. In the case of a strained semiconductor the ϵ_{ij} are the components of the (symmetric) strain tensor for cubic materials [57]. The CB edge E_c and the absolute, average VB energy E_v' (defined from the model-solid theory [58]), as well as the Kane parameters A', L', M, N', and P, and the strain coefficients l, m, and n are calculated by [33, 48]

$$\begin{aligned} E_c &= E_v + V_{\mathrm{ext}} + E_0 \\ E_v' &= E_v + V_{\mathrm{ext}} - \frac{\Delta_0}{3} \\ A' &= \frac{\hbar^2}{2m_0}\left(\frac{1}{m_e} - \frac{E_p}{E_0}\,\frac{E_0 + 2\Delta_0/3}{E_0+\Delta_0}\right) \\ P &= \sqrt{\frac{\hbar^2}{2m_0}E_p} \\ L' &= \frac{P^2}{E_0} - \frac{\hbar^2}{2\,m_0}(1+\gamma_1+4\gamma_2) \\ M &= -\frac{\hbar^2}{2\,m_0}(1+\gamma_1-2\gamma_2) \\ N' &= \frac{P^2}{E_0} - \frac{3\,\hbar^2}{m_0}\gamma_3 \\ l &= 2b_v + a_c - a_g \\ m &= a_c - a_g - b_v \\ n &= \sqrt{3}\,d_v \end{aligned}$$

so that the eight-band **k·p** Hamiltonian for strained bulk is parametrized by

- the fundamental band gap E_0 ,
- the SO energy Δ_0 ,
- the optical matrix parameter E_p ,
- the VB edge E_v ,
- the relative Γ-point CB mass m_e ,
- the three Luttinger parameters γ_1 , γ_2 , and γ_3 ,
- the Kane parameter B,
- the hydrostatic CB deformation potential a_c ,
- the hydrostatic band gap deformation potential a_g ,
- the uniaxial ([100] direction) VB deformation potential b_v ,
- the uniaxial ([111] direction) VB deformation potential d_v ,
- the parameter b' coupling the CB edge to shear strain,
- and an optional scalar potential V_{ext} describing an electric field resulting, e. g., from a built-in voltage in a p-n-junction, an externally applied voltage, or a piezoelectric charging. Only the latter case is considered in this work.

The hydrostatic gap deformation potential is related to the hydrostatic CB and VB deformation potentials by $a_g = a_c - a_v$. Using the relation

$$E_p \;=\; \frac{3\,E_0\,(E_0+\Delta_0)}{\Delta_0}\left(\gamma_1 - \frac{1}{m_{\text{SO}}}\right) \qquad ,$$

the Kane parameters A' and P can alternatively be calculated using m_{SO} instead of E_p [33].

The cubic strain tensor [57], represented by its independent components ϵ_{xx}, ϵ_{xy}, ϵ_{xz} , ϵ_{yy} , ϵ_{yz} , and ϵ_{zz} , is defined for infinitesimal displacements, and its inclusion to the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian by means of linear deformation potential theory [49, 56] is based on the assumption of small strain, like $|\epsilon_{ij}| < 0.05$.

In calculations for bulk semiconductors, the VB edge energy E_v can have any value as it only defines the energy gauge. In calculations of heterostructures, however, E_v determines the CB and VB offsets at the heterojunctions.

2.2 Heterostructures

In case of a heterostructure the above mentioned material parameters become spatially dependent. If the heterostructure, in addition, is inhomogeneously strained also the strain tensor is a function of the spatial position. The six Kane parameters A', B, L', M, N', and P, and the six strain components appear in the matrix G in products together with the wave vector $\mathbf{k}$. Since the spatial position operator and the wave number operator do not commute, G – and thus the Hamiltonian $\hat{H}$ – is not generically Hermitian for heterostructures.

In order to retain the Hermiticity of the Hamiltonian it is necessary to replace any such product of wave vector components k_i with a Kane parameter or a strain component (both called Q below) by respective symmetrized products:

$$\begin{aligned} Qk_i &\rightarrow (Qk_i + k_iQ)/2 \qquad , \\ Qk_ik_j &\rightarrow (k_iQk_j + k_jQk_i)/2 \qquad , \\ i,j &= x,y,z \qquad . \end{aligned}$$

Assuming identical Γ-point Bloch functions in all constituent materials of the heterostructure the symmetrization implies probability flux conservation at all heterointerfaces, regardless of their orientation [59].

Explicitly, inhomogeneous strain is treated in an adiabatic manner by this approach, but implicitly, part of the inhomogeneity (expressed by $\nabla\hat{\epsilon}$ terms) effects on the bandstructure is included to the Hamiltonian due to the symmetrization [34].

2.3 Material parameters

The eight-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian provides a rather detailed electronic model of a semiconductor as it employs 15 parameters to describe the Brillouin zone (BZ) center bandstructure in presence of a given strain situation, assuming that these parameters are not strain-dependent theirselves. A strain dependence of, e. g., the SO interaction [60] is neglected. Some of these parameters, as E_0 , Δ_0 , and m_e , are known to large accuracy from experiments and are available for all technologically important semiconductors from the literature. Others, like the Luttinger parameters, E_p , a_g , and E_v' (i. e., band offsets), are more difficult to determine experimentally since their measurement is mostly indirect and complicated by trade-offs with other uncertain quantities. Hence, often a variety of values exists for those parameters, necessitating a judicious selection. There are also parameters for which experimental values either are not available or show such a large spread that they provide poor guidance. For those, sometimes calculated values have to be assumed, or they are typically neglected due to the lack of any reasonable estimate, as, e. g., B and b'.

As an example for the difficulties connected to the choice of values for the linear deformation potentials, Fig. 2.4 displays various literature values (both experimental and calculated) for the hydrostatic deformation potentials a_c , a_g , and a_v of GaAs, compiled from Ref. [45, 61, 62]. While the magnitude of the gap deformation potential a_g is relatively well determined, the distribution of the hydrostatic band-gap shift over the CB and VB is considerably less easily examined. It is not even evident whether the VB shifts to higher or lower energies under tensile strain, a question which is relevant for the VB band offset at strained heterojunctions. In the present work, average values have been employed, denoted by the black disk in Fig. 2.4.

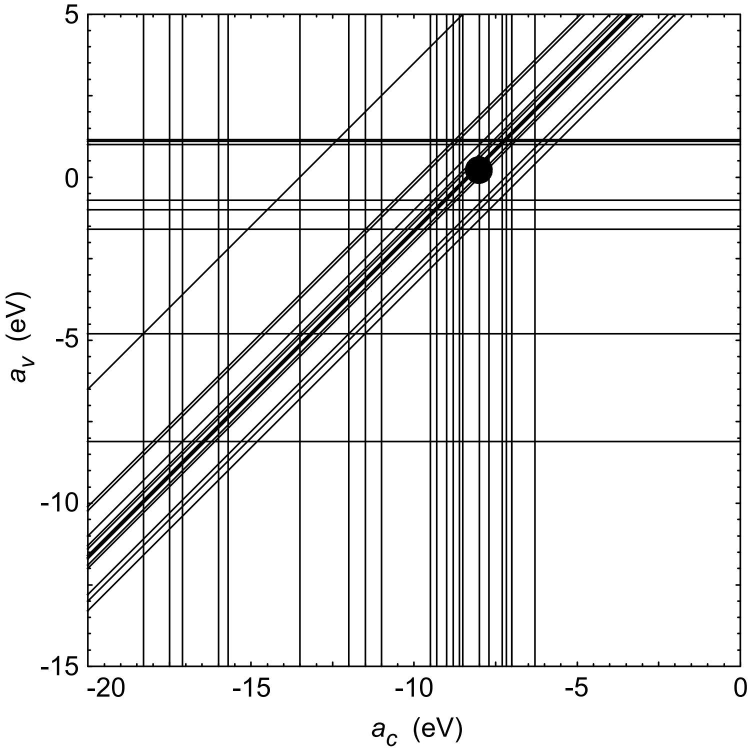


Figure 2.4:

Literature values [45,61,62] for the hydrostatic deformation potentials a_c (vertical lines), a_g (diagonal lines), and a_v (horizontal lines) of GaAs. The black disk indicates the triple of values used in this work.

In addition to the uncertainties regarding the deformation potentials used in linear deformation potential theory [56], the strain dependence of the band edges is believed to be (weakly) non-linear, as suggested by experiments [63] and second-principles calculations [38,64,65]. For the modeling of moderately strained, pseudomorphic QWRs such effects may safely be neglected. In contrast, the strain encountered in Stranski-Krastanow grown QDs reaches a magnitude where the non-linear strain dependence of the band edges is, in principle, mandatory to consider. The inclusion of these non-linearities to actual calculations is, however, hampered by the persisting uncertainty whether a *super*linearity or a *sub*linearity has to be assumed, see Fig. 2.5. As a "compromise", the inclusion of the non-linearity is suspended in this work, in expectation of unambiguous answers to this question.

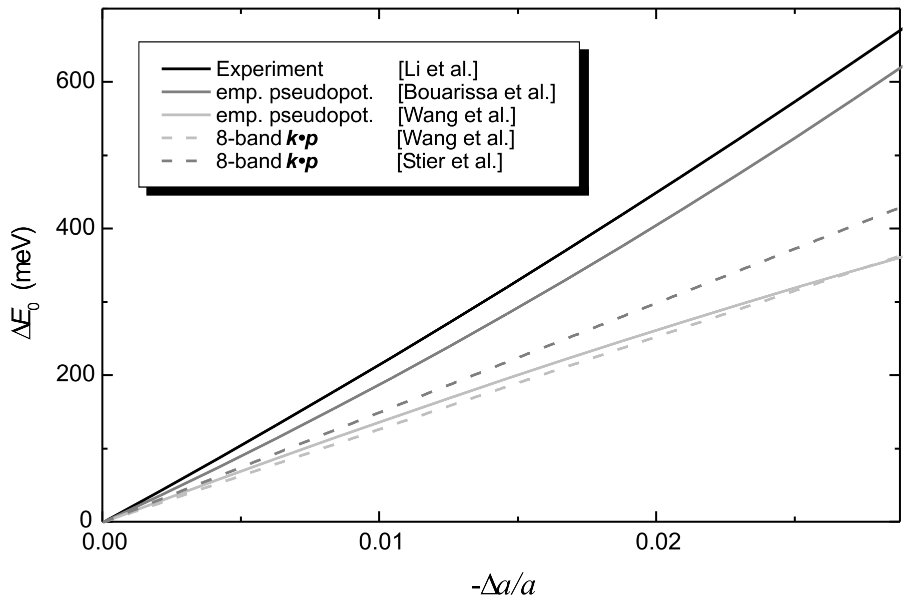


Figure 2.5:

Hydrostatic strain dependence of the fundamental gap E_0 of InAs according to different calculations (gray lines) and one experiment (black line) made on monolayer thick InAs quantum wells in GaAs [63]. The change of E_0 is plotted vs. the change of the lattice constant. The solid gray lines show EP calculations [38,64], the dashed gray lines are calculated from Eq. (2.1) using the material parameters given in Ref. [38] and [31], respectively.

Another important, though insecure, material property is the VB offset at a strained heterojunction. In the present approach, the problem is partitioned into (i) the determination of an intrinsic, absolute band offset between bulk materials not connected to each other, and (ii) the modification of this offset due to strain, when matching the respective lattices. The second part requires "only" the knowledge of the strain on both sides of the interface, as well as the deformation potentials of the respective materials. Effects resulting from interface dipoles, or a possible dependency on the interface orientation or the order in which adjacent layers have been grown, are thereby neglected. The first part (i) is based on the notion that the VB edge positions of different semiconductors are defined on one and the same energy scale, thus presenting an intrinsic property the bulk material [58, 66, 67].

These intrinsic band offsets have been derived from a variety of experiments for the materials relevant to this work. The scatter of values for the (almost) lattice-matched AlGaAs material system [62, 66, 68] is not as large as that obtained for the InGaAs/InP material system shown in Fig. 2.6.

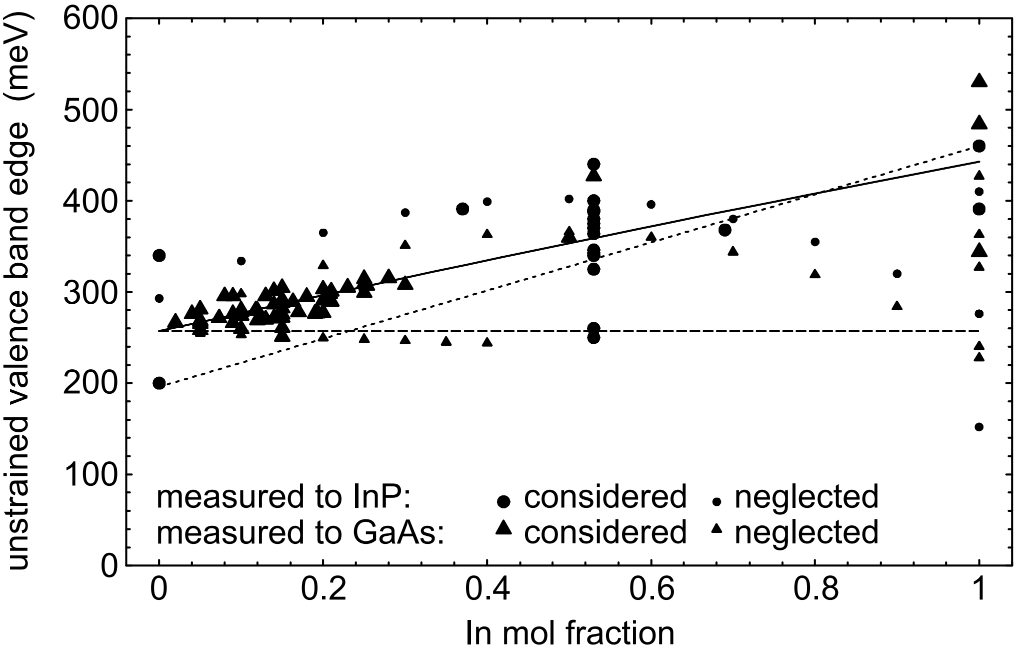


Figure 2.6:

Energy of the unstrained (intrinsic) VB edge of $In_xGa_{1-x}As$ relative to that of InP. Triangles denote experimentally derived values for InGaAs/GaAs heterojunctions, circles refer to measurements on InGaAs/InP heterojunctions. The solid line is a quadratic fit to all values represented by large symbols, neglecting values represented by small symbols. The horizontal, long-dashed line indicates the GaAs/InP VB offset. The short-dashed line is a theoretical prediction from the model-solid theory [58]. The data points are compiled from Ref. [45, 62, 69].

The solid line in Fig. 2.6 indicates the VB alignment assumed in this work, obtained by a quadratic polynomial fit to selected experimental values (represented by the large triangles and circles). The resulting VB energies of InGaAs on InP are shown in Fig. 2.7, and those of InGaAs or AlGaAs on GaAs in Fig. 2.8.

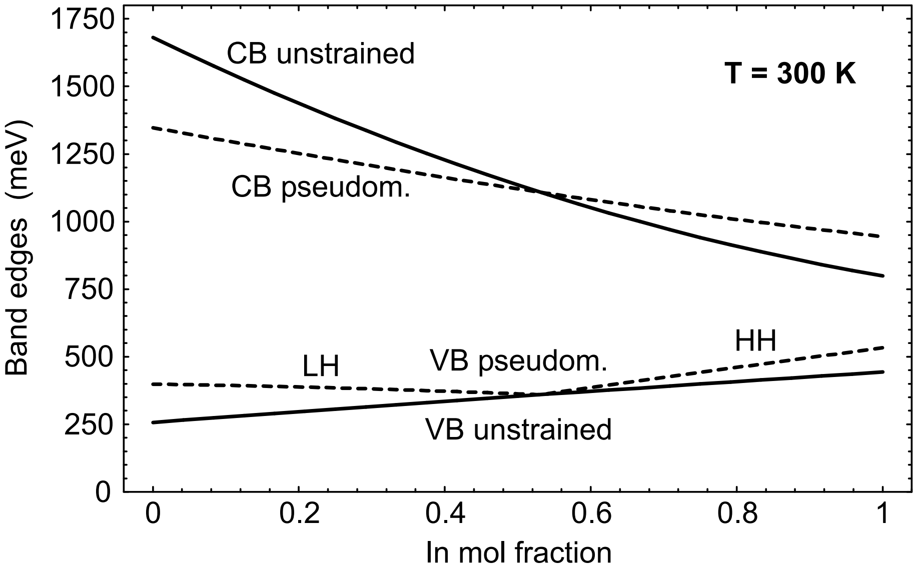


Figure 2.7:

Band edge energies of InGaAs on InP at room temperature, as assumed in this work.

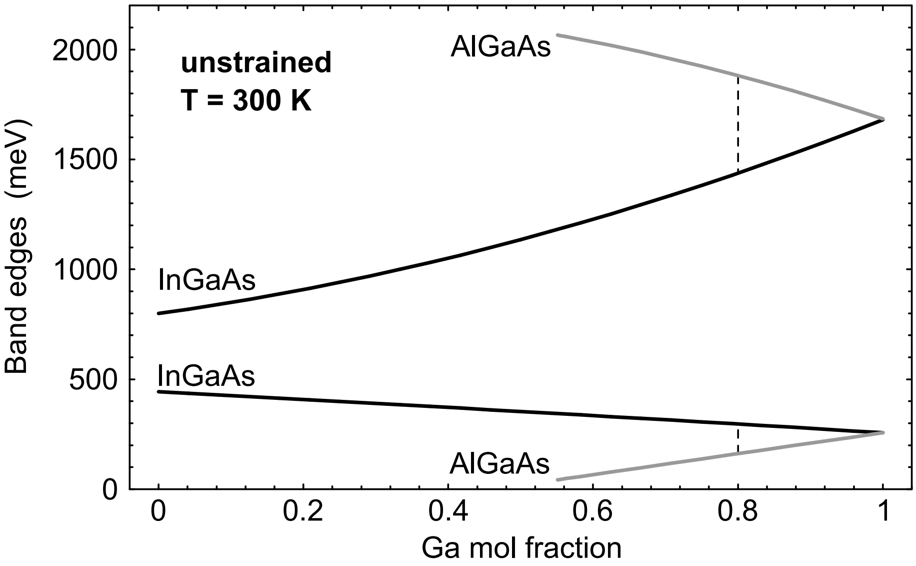


Figure 2.8:

Unstrained (intrinsic) band edge energies of InGaAs (black lines) and AlGaAs (gray lines) at room temperature with respect to InP, as assumed in this work. The vertical, dashed lines indicate the assumed band alignment at the heterointerface of the QWR studied in Section 7.1.

Chapter 3

Quantum-Confined Heterostructures

The suitability of the eight-band **k**·**p** model for the modeling of strongly confining and inhomogeneously strained nanostructures decreases with decreasing QD or QWR sizes due to inherent, conceptional approximations [31, 38, 70]. The major limitations result from

- the validity of the **k**·**p** scheme being restricted to the immediate vicinity of the BZ center Γ,
- the relatively small number of Bloch functions used for the expansion of the wave functions (eight in the eight-band **k**·**p** model),
- the assumption of identical Bloch functions throughout the heterostructure (regardless of material and strain variations),
- and the seeming arbitrariness of the envelope matching conditions at heterointerfaces.

Similar limitations do not exist in microscopic approaches like empirical pseudopotential (EP) theory, which have potentially greater accuracy. To improve the accuracy of the eight-band **k**·**p** model applied to small, highly strained QDs, a generalized concept for using the eight-band **k**·**p** scheme in electronic structure calculations of QDs is proposed, which overcomes part of the above theoretical problems: In practice, it is sufficient to fit the **k**·**p** parameters of the constituent materials to the known bulk bandstructures within an appropriately large BZ segment around the Γ point. Thus the eight-band $\mathbf{k} \cdot \mathbf{p}$ model, though originally derived for an infinitesimal **k**-space region around the Γ point [48] and infinitesimal, homogeneous strain [56], yields accurate predictions even for QD heterostructures with large quantization energies, strong inhomogeneous strain, and piezoelectric effects. After such an extension of the valid **k**-range the remaining **k·p** approximations appear hardly significant, as shown in this chapter.

The concept is demonstrated for pyramidal InAs/GaAs QDs by a comparison with atomistic calculations reported in Ref. [38]. The single-particle energy levels, wave functions, Coulomb energies, and momentum matrix elements are calculated and compared to the results of prevailing EP calculations for such QDs [36–39,71]. Prior to the calculation of confined levels, the **k·p** bulk bandstructure parameters are fitted to reproduce the bulk bandstructures of InAs and GaAs, as derived by the pseudopotential theory, to the largest possible extent. An assessment of the eight-band **k·p** model by comparison with experimental data is limited due to the insufficient accuracy of available structural and optical data, as discussed in Section 6.3.

The choice of InAs/GaAs QDs serves here as an example only: To model QD structures formed by other materials one has, in the same manner as described below, to determine the **k·p** parameters from fits to the respective bulk bandstructures as calculated from, e. g., first principles or (semi)empirical pseudopotential theory, being available for all common semiconductors from the literature.

3.1 The model quantum dot

As a test bed for the explanation and demonstration of the proposed procedure, a pyramidal InAs QD with {101} side facets and a base length of 20 lattice constants (11.3 nm) on top of a one monolayer (ML) thick (001) wetting layer (WL) of InAs is considered, which is coherently embedded in GaAs and thus under pseudomorphic strain. The assumed temperature is 0 K. The electronic structure of this model QD was calculated twice by the EP method, first neglecting the SO interaction in the constituing bulk materials [37], and then including it [38]. The single-particle energy levels obtained from the two EP calculations are shown in Fig. 3.1(a,b), respectively. The wave functions belonging to term scheme (a) (neglecting SO interaction) are shown in Figs. 1 and 2 of Ref. [37]. The **k·p** calculations in this chapter are going to be compared to term scheme (b) (including SO interaction), the corresponding EP calculated wave functions are shown in Fig. 3 of Ref. [38] and Fig. 2 of Ref. [71].

3.1.1 Strain effects

The strain distribution in the QD and surrounding barrier has a dominant influence on the electronic structure and constitutes, together with the deformation

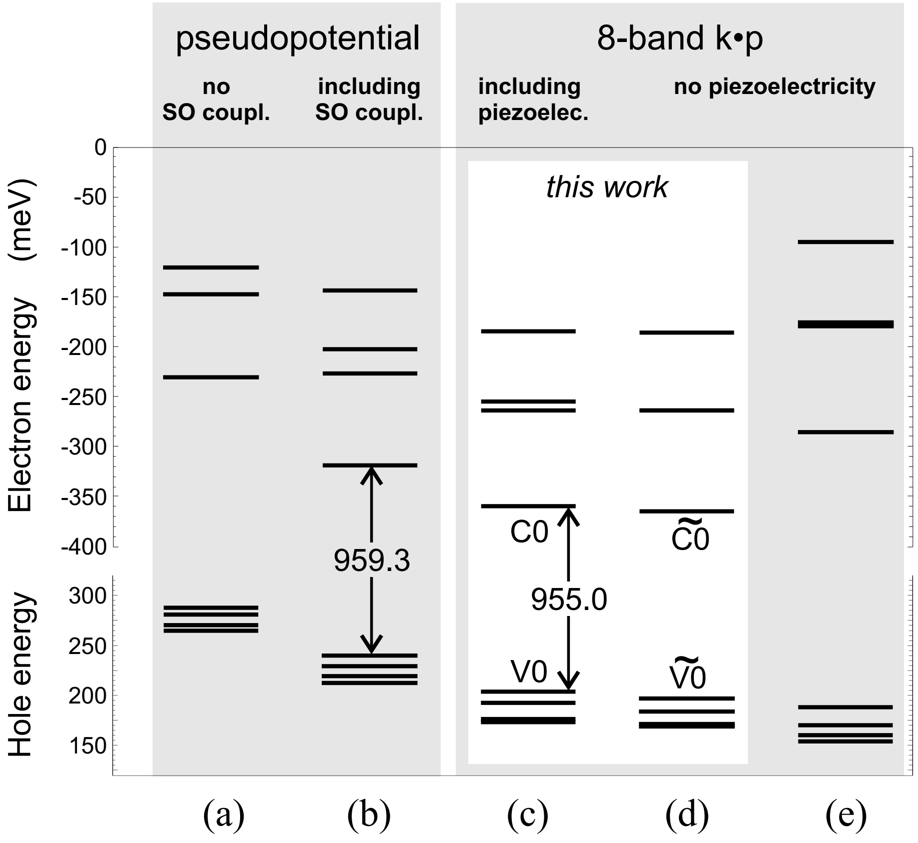


Figure 3.1:

Single-particle energy levels in the model QD assumed in this chapter, according to different calculations. The electron (hole) energies refer to the CB (VB) edge of the GaAs barrier far away from the dot (unstrained). (a): EP calculation neglecting SO coupling in the bulk materials, from Ref. [37]. (b): EP calculation including SO coupling and compared to the present calculations, from Ref. [38]. (c): Eight-band **k·p** calculation using the parameters from Table 3.1 and including the piezoelectric effect. (d): Eight-band **k·p** calculation using the parameters from Table 3.1 and neglecting the piezoelectric effect. (e): Eight-band **k·p** calculation from Ref. [71], neglecting the piezoelectric effect.

potentials [56], critical input data for the **k·p** model. The strain calculation[1] in this chapter follows Ref. [38], using Keating's valence force field model (Eq. (4.4) in Section 4.1.2) with the elastic parameters from Ref. [26] and a $69a \times 69a \times 69a$ supercell with periodic boundary conditions (a is the GaAs lattice constant). By different conventions, the [110] direction in Ref. [38, 71] is the $[1\bar{1}0]$ direction here and *vice versa*. The strain distribution possesses C_{2v} symmetry due to the atomistic property of the strain model which takes into account the inequivalence between the [110] and [1$\bar{1}$0] directions in the zincblende lattice (see Section 4.1.2).

---

[1] The strain distribution has been calculated by Marius Grundmann.

The strain dependence of the bulk band edges is nonlinear according to the EP calculation [65]: The fundamental bandgap E_0 of (mainly) hydrostatically strained bulk InAs scales sublinearly with the strain $\hat{\epsilon}$. It follows that linear deformation potential theory [56] would overestimate E_0 at large strain ($\mathrm{Tr}\,\hat{\epsilon} \approx -0.08$) by $\sim 50\,\mathrm{meV}$ if the values for the Γ-point deformation potentials were calculated within the small-strain limit ($|\mathrm{Tr}\,\hat{\epsilon}| \leq 0.01$). To avoid such inaccuracies, hydrostatic non-linearities of E_0 can be incorporated in a straight-forward way into the **k·p** model using, for instance, experimentally known values of the respective coefficients [63, 69, 72].

In the given case, however, the linear deformation potentials previously derived from the EP calculation and listed in Table II of Ref. [38] fit well an average high-strain regime:[2] The local CB and VB edges at a position $\sim 3.8\,\mathrm{nm}$ below the pyramid's tip as calculated by eight-band **k·p** theory agree within 10 meV with the EP calculated band edges shown in Fig. 5 of Ref. [38]. In view of this match, linear deformation potential theory with the deformation potentials from Ref. [38] are used for the intended comparison. Due to the small, but inevitable, deviations between the confinement potentials in the EP and **k·p** calculations, any deviation $\leq 10\,\mathrm{meV}$ between the energy levels calculated by the respective models deserves no discussion.

3.1.2 Piezoelectricity

The pseudopotentials [38] describe the reaction of the bulk bandstructure on the elementary lattice deformations, i. e., due to hydrostatic pressure or uniaxial stress along [100] or along [111]. However, due to the neglection of long-range electric fields they do not correctly model the piezoelectric effect [73] arising for uniaxial stress along [111] from the partial cancellation of the ionic polarization and the compensating valence hull deformation [74–78]. The piezoelectric effect can, on the other hand, be included into the **k·p** calculation in a straight-forward manner described in Section 4.2, by using experimentally determined piezoelectric moduli and adding an electrostatic potential to the Hamiltonian.

Thus, regarding piezoelectricity the comparability of the present **k·p** calculations with the previous EP calculations is limited. Nonetheless, self-organized QDs are formed by a strain driven growth mechanism [19] and thus are always strained, while the constituent materials, like In(Ga)As/Ga(Al,P)As or Cd(Zn,S)Se/Zn(S)Se, are piezoelectric. Hence, the inclusion of piezoelectricity improves the description of reality. Second, the comparison made below between **k·p** calculations including piezoelectricity and the EP calculations neglecting it shows that the effects originating from the atomic structure anisotropy (ASA) [38] between the [110] and [$1\bar{1}0$] directions and from piezoelectricity are qualitatively and quantitatively similar – although piezoelectricity is related to strain while the ASA is not. Common to both mechanisms is the symmetry reduction of the strained unit cell from C_{4v} to C_{2v}.

---

[2] This is also visualized by the intersection of the two light gray lines in Fig. 2.5.

3.1.3 Bandstructure

So far, the concern has been to ensure good agreement between the potential energies of bound states in both calculations by realizing equivalent confinements. Equally important is to assume possibly equivalent bulk dispersion relations in order to create comparable starting points for the calculation of the kinetic energies by both methods. Therefore, the eight-band **k**·**p** bandstructure parameters are selected here as to provide the best available agreement with the bulk bandstructures calculated from the pseudopotentials within the largest possible region around the Γ-point. This intention advances beyond the calculation of the **k**·**p** parameters from the zone-center wavefunctions according to their definitions by perturbation theory [48, 79].

This generalized notion of the **k**·**p** method aims at realizing the largest possible validity range of the **k**·**p** calculated bandstructure. Given a different QD structure formed by different materials, the prevailing bulk bandstructures of these materials, as known from first- or second-principles calculations or measurements, are submitted to a parameter fit procedure [79, 80] yielding the optimal **k**·**p** bandstructure parameters according to the above criterion. Thereby, a larger range of validity of the **k**·**p** bandstructure on both the **k**- and the energy axis is made available than by most of the usual **k**·**p** parameter values defined for the immediate vicinity of Γ only. The approach is applicable to any direct semiconductor whose bandstructure has been calculated or measured before. **k**·**p** parameters for alloys can be derived accordingly or by appropriate interpolation.

Fig. 3.2(a) shows the present fit of the eight-band **k**·**p** bandstructure to the EP calculated bandstructure of bulk InAs [81]. The **k**·**p** bandstructure is defined by the Γ-point CB mass m_e, the Luttinger parameters γ_1, γ_2, γ_3, the Kane parameter B, the optical matrix parameter $E_p = 2m_0P^2/\hbar^2$, the SO energy Δ_0, and the gap E_0 [48]. Except for E_0 (fixed) these parameters were fitted to reproduce the EP bandstructure for $|\mathbf{k}| \leq 1.8\,\mathrm{nm}^{-1}$. In the same manner the GaAs parameters were fitted for $|\mathbf{k}| \leq 1.5\,\mathrm{nm}^{-1}$ to an EP bandstructure compiled from Ref. [82,83]. The resulting parameter values are summarized in Table 3.1.

The task to fit a given bandstructure within a possibly large region of the BZ may force a departure from the usual literature values[3] for the **k**·**p** parameters.

---

[3] See, e. g., Ref. [45].

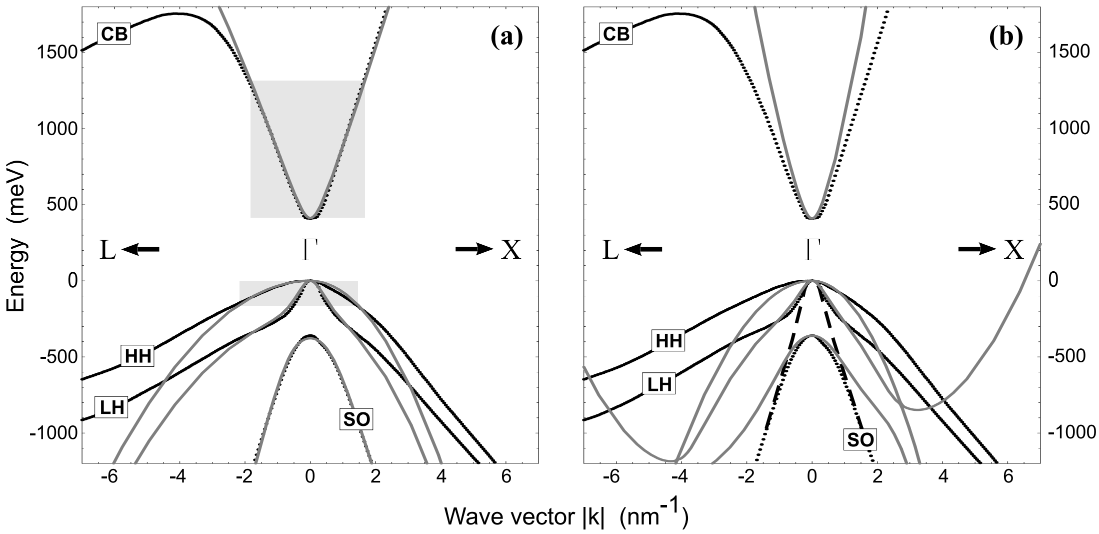


Figure 3.2:

Bandstructure of bulk InAs in the vicinity of $\mathbf{k} = \mathbf{0}$ at 0 K, showing the CB, heavy hole (HH), light hole (LH), and SO bands. Black narrow-dotted lines: EP calculation from Ref. [81]. Gray lines: Eight-band $\mathbf{k}{\cdot}\mathbf{p}$ calculation using the parameters (a) from Table 3.1 or (b) from Ref. [71], respectively. (a): Gray rectangles show the range of good agreement between the $\mathbf{k}\cdot\mathbf{p}$ and EP bandstructures. (b) The black dashed line sketches the merging of the LH and SO bands to one band, as obtained if the SO interaction is neglected in the EP calculation.

While the latter values describe the bandstructure within the immediate vicinity of the Γ-point, they generally yield too steep CB and VB dispersions at larger $|\mathbf{k}|$ values. Calculations of strongly quantized states using those parameters [45] will therefore systematically overestimate the kinetic energies, i. e. the quantization. In contrast, the values in Table 3.1 provide an improved description of the (EP) bandstructure throughout a larger part of the BZ and a larger energetic range, as marked by the gray shading in Fig. 3.2(a), sacrificing accuracy directly at the Γ-point. Since significantly quantized states lay energetically well apart from the band edges, the reduced accuracy *at* Γ is negligible for the electronic properties of QD structures.

$B = 0$ in Table 3.1 is also a fit result, meaning that the bandstructures could, as well, belong to a diamond structure semiconductor. The inversion asymmetry of the bulk lattices is not important *throughout* the entire selected fitting range. If the fit is restricted to a much smaller $|\mathbf{k}|$ range, B values in the order of 100-200 meV nm^2 are found for InAs. However, fixing $B = 100\,\text{meV}\,\text{nm}^2$ has little influence on the bound QD states so that $B = 0$ is assumed in the calculations here.

		Unit	Value for InAs	Value for GaAs
lattice constant		nm	0.60553	0.56533
piezoelectric modulus	e_{14}	$\mathrm{C\,m^{-2}}$	0.045	0.16
static dielectric constant	$\epsilon(0)$	—	14.6	13.18
SO energy	Δ_0	meV	378	261
optical matrix parameter	E_p	eV	12.12	17.52
Kane parameter	B	$\mathrm{meV\,nm^2}$	0	0
CB effective mass (Γ)	m_e	m_0	0.0406	0.0904
Luttinger parameter	γ_1		10.08	4.47
Luttinger parameter	γ_2		4.57	1.52
Luttinger parameter	γ_3		5.11	2.16
CB-VB coupling by strain	b'	meV	0	0

Table 3.1:

Material parameters used in the $\mathbf{k}\cdot\mathbf{p}$ calculations of this chapter. For all other properties the same values as in Ref. [38, 71] were taken.

To demonstrate the extended predictive power of the eight-band $\mathbf{k}\cdot\mathbf{p}$ model resulting from the explained parameter selection procedure, the 0D energy level structure in the model QDs is calculated next. Thereafter, the results are compared (i) to the prevailing EP calculations [38], demonstrating the high accuracy of the eight-band $\mathbf{k}\cdot\mathbf{p}$ calculation, and (ii) to previous eight-band $\mathbf{k}\cdot\mathbf{p}$ calculations [71] using "classically" defined parameters, demonstrating the importance of the selection procedure and, thus, the related enhancement of accuracy.

The bound states in the model QD are calculated as described in Section 4.3, using the strain-dependent eight-band $\mathbf{k}{\cdot}\mathbf{p}$ Hamiltonian Eq. (2.1). By this approach the strain is treated adiabatically, whereby strain inhomogeneity (expressed by terms of $\nabla\hat{\epsilon}$) is implicitly included [34] in a manner analogous to the treatment of image charge effects in Eq. (4.12), i. e., by symmetrization [33] of $\epsilon_{ab}k_c$ products in Eq. (2.2). For the calculation of the term scheme Fig. 3.1(c) the piezoelectric effect was included according to Section 4.2, using the piezoelectric moduli from Table 3.1. Term scheme Fig. 3.1(d) was calculated without piezoelectricity. The Hamiltonian was discretized by finite differences (see Chapter 8) on a cubic mesh with $72\times72\times60$ voxels and 0.2827 nm spatial resolution (half the lattice constant a of GaAs) and diagonalized by the algorithm described in Section 9.3.

From the single-particle eigenstates the momentum matrix elements of electron-hole recombinations for given linear polarization are calculated using Eq. (4.21) and Eq. (4.24). The (integral) transition matrix element is obtained by integrating Eq. (4.21) over all polarization directions. The Coulomb matrix elements between electrons and holes (denoted by C in Table 5.3) are calculated using the bulk dielectric constants from Table 3.1 and including image charge effects.

3.2 Predicted quantum dot properties

The discussion is restricted to four bound CB and VB states, each, although additional bound electron and hole states are obtained in the model QD. The states are labeled by C for CB or V for VB plus an integer numbering the excitation levels, 0 denoting the groundstates.

The projections of a bound-state wave function to the four Γ-point bulk bands, *s* (CB), heavy hole (HH), light hole (LH), and SO – as displayed in Fig. 3.3(e-g) – show that the total probability density looks like possessing a well-defined parity if one of the projections has an intensity above 80%, i. e., if the contributions from remote bands sum up to less than 20%. Otherwise the band mixing becomes directly visible in the probability density by a smearing-out of nodal surfaces or an entirely non-elementary shape.

3.2.1 Neglection of the piezoelectric effect

The calculation neglecting piezoelectricity corresponds one-to-one to the EP calculation [38], as well as to another eight-band $\mathbf{k}\cdot\mathbf{p}$ calculation [71]. The present results differ from those in Ref. [71] in some respects detailed below. The bound states calculated without piezoelectricity are denoted by a tilde (˜) to distinguish them from those calculated including piezoelectricity.

Wave function compositions and symmetries

The probability densities are plotted in Fig. 3.3(a,b) and Fig. 3.4(a,c). The CB states exhibit a recognizable parity since the remote contributions from the VB amount to 12-18% only. *Vice versa*, the *s*-type Bloch function contribution to the VB states is less than 1.4%. The least mixed state is the hole groundstate $\tilde{V0}$ which is even more HH-like (88.4%) than the electron groundstate $\tilde{C0}$ is *s*-like (87.9%). From $\tilde{V0}$ to $\tilde{V3}$ the LH projection increases from 9% to 20% while the SO projection stays between 2.2% and 4.7%. Thus (i) the calculation of the CB states requires the inclusion of *p*-type Bloch functions whereas (ii) the *s*-type Bloch functions are not so important for the VB states.

The weak C_{2v} symmetry of the strain distribution is mainly reflected by the order of the excited electron states $\tilde{C1}$ and $\tilde{C2}$. The spatial band edge profiles shown in the top row of Fig. 3.4 favour an electron expansion along the $[1\bar{1}0]$

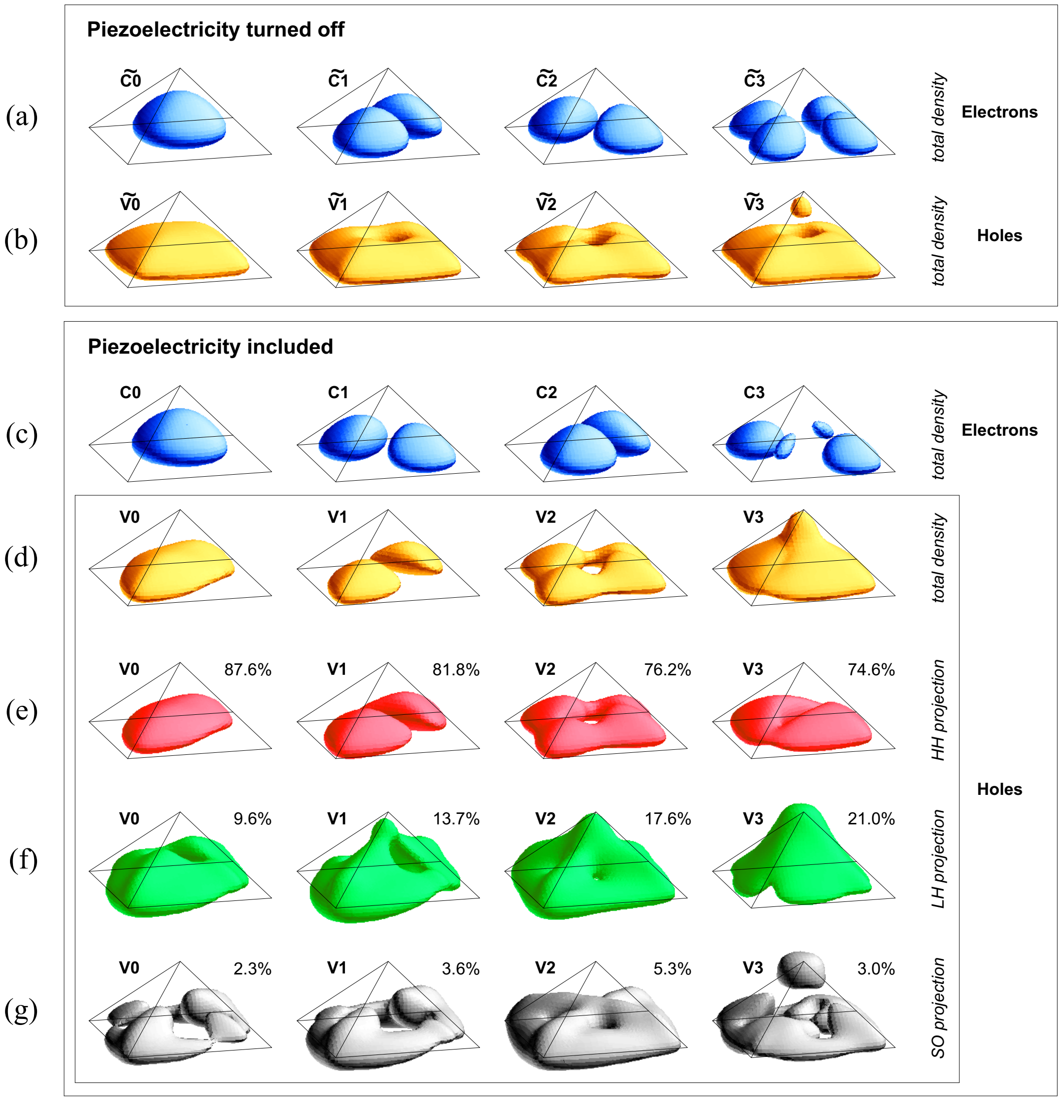

Piezoelectricity turned off
(a)
C̃0
C̃1
C̃2
C̃3
total density
Electrons
(b)
Ṽ0
Ṽ1
Ṽ2
Ṽ3
total density
Holes
Piezoelectricity included
(c)
C0
C1
C2
C3
total density
Electrons
(d)
V0
V1
V2
V3
total density
(e)
V0 87.6%
V1 81.8%
V2 76.2%
V3 74.6%
HH projection
Holes
(f)
V0 9.6%
V1 13.7%
V2 17.6%
V3 21.0%
LH projection
(g)
V0 2.3%
V1 3.6%
V2 5.3%
V3 3.0%
SO projection

Figure 3.3:

Isosurface plots (25% of the maximum value) of total probability densities (a-d) and VB projection intensities (e-g) of bound electron (a,c) and hole (b,d-g) states in the model QD, calculated by eight-band $\mathbf{k}\cdot\mathbf{p}$ theory using the parameters from Table 3.1 and including (c,d) or neglecting (a,b) the piezoelectric effect, respectively. The surface triangulation shows the spatial resolution of the calculation. The percent numbers in rows (e-g) are the integral intensities of the projections to the HH, LH, or SO bulk bands, respectively, and the isosurfaces show their shapes. For each state $V0\ldots V3$ the difference to 100% is the integral intensity of the s-type Bloch function projection which is not shown here.

direction over one along [110]. The corresponding level separation of 0.3 meV is, however, negligible since the ASA enters the $\mathbf{k}\cdot\mathbf{p}$ calculation only via the

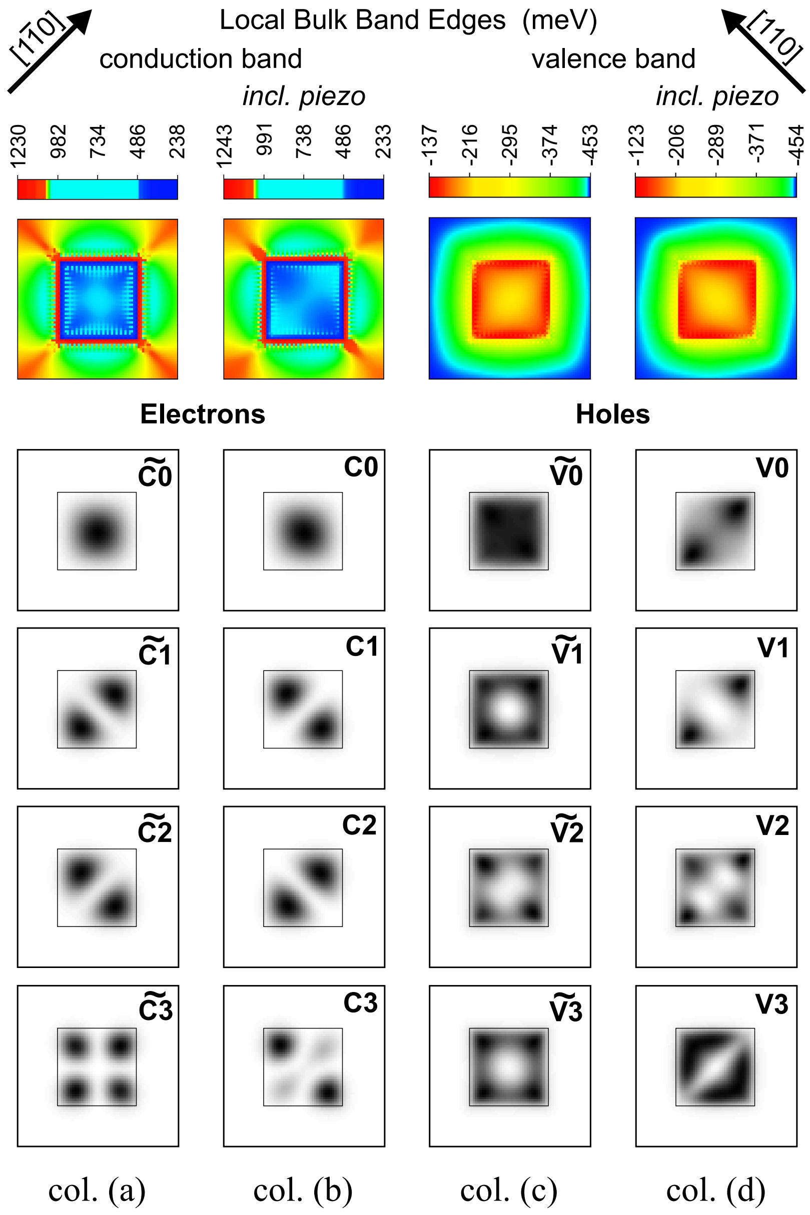


Figure 3.4:

Top row: CB (cols. a,b) and VB (cols. c,d) edge profiles in a (001) cross section plane 1.0 nm above the WL, pixelwise calculated from the strain dependent eight-band **k**·**p** Hamiltonian Eq. (2.1). The energies refer to the CB edge of unstrained InAs (being 1060 meV below the unstrained GaAs CB edge). For better contrast a non-linear color scale is used. Due to piezoelectricity the profiles are C_{2v} symmetric (cols. b,d) while at neglection of piezoelectricity the symmetry is almost C_{4v} (cols. a,c). Other rows: Corresponding probability densities of bound states along the same section plane on a linear gray scale, calulated by eight-band **k**·**p** theory using the parameters from Table 3.1.

strain distribution. In the EP calculation the ASA has an additional, strain-independent electronic effect leading to 24 meV separation of $\tilde{C}1$ and $\tilde{C}2$. The VB groundstate $\tilde{V}0$ shows a slight preference for the [110] direction but the excited VB states prefer $[1\bar{1}0]$, like the CB states.

Spectral properties

The energy levels are shown in Fig. 3.1(d). The eight-band **k·p** CB (VB) levels are 47 ± 9 (45 ± 2) meV lower than those from the EP calculation while the ground-state transition energy $(\tilde{C}1 - \tilde{V}1)$ agrees with the EP value within 2.5 meV. Hence, the eight-band $\mathbf{k}\cdot\mathbf{p}$ model predicts accurate kinetic (quantization) and transition energies. Rigid spectral shifts of the magnitude of ~ 46 meV occur, e. g., when the band offsets at the heterojunction are altered by a similar amount. Other potential reasons for a rigid shift are remaining deviations in the strain dependence of the band edge profiles, or in the GaAs bandstructure. The shift is considered insignificant here because (i) it is rigid, i. e., the energy differences from the present calculation are very similar to the corresponding EP values, and (ii) the spectral shift between the two EP calculations with [38] and without [37] inclusion of the SO interaction is even larger.

The good agreement between the EP calculation and the present eight-band $\mathbf{k}\cdot\mathbf{p}$ calculation has to be contrasted with an earlier comparison stating larger discrepancies [71]. The selection procedure for the **k·p** bandstructure parameters proposed here is decisive, as detailed in the following.

Band mixing effects

The energy levels from the eight-band $\mathbf{k}\cdot\mathbf{p}$ calculation in Ref. [71] are shown in term scheme Fig. 3.1(e). The CB (VB) levels are in average 86 (12) meV higher (lower) and the ground-state transition energy is 88 meV larger than the values obtained here. The reason for this discrepancy is that in Ref. [71] different bandstructure parameters were used which, for InAs, produce the bandstructure shown in Fig. 3.2(b). This bandstructure does not approximate the EP bandstructure within a sufficiently large region around Γ. In the classical sense, the parameters used in Ref. [71] yield a correct **k·p** representation, however, they do not meet the high accuracy demands for QD calculations: Particularly the CB and HH masses are too small, seen over a larger energy range, thus overestimating the kinetic energies (overconfinement).

The somewhat too large LH-HH mass ratio in the calculation of Ref. [71] also accounts for a cross-over of two excited VB states as compared to the EP calculation. Fig. 3.3(b) shows a pronounced leakage of the $\tilde{V}3$ wavefunction towards the pyramid tip due to the large LH component (20%): Following the respective band edge profiles HH wave functions tend to the pyramid bottom while LH ones are attracted to the tip [39]. In agreement with the EP calculation the $\tilde{V}3$ state calculated here has the strongest LH component, realizing the most groundstate-like projection to the LH band. In the $\mathbf{k}\cdot\mathbf{p}$ calculation of Ref. [71], however, this feature appeared already in $\tilde{V}2$ due to the incorrect mass ratio which overestimates the LH projection. Also the somewhat larger VB level separations obtained there result from the smaller average VB mass.

The opposite case of an unnaturally small LH-HH mass ratio for large $\mathbf{k}$ values is encountered if the SO interaction is neglected, as in an earlier EP calculation [37], see Fig. 3.1(a). The corresponding bulk bandstructure is sketched by the dashed line in Fig. 3.2(b) which shows the reduction of the Γ point VB structure to two bands by merging the former LH and SO bands. The new "LH" band is so much lighter than the HH band that their mixing is largely suppressed. Eventually, this results in an almost single-band representation involving only the HH band. Then, bound VB states have well developed parities and smaller level separations, the LH-related tendency to propagate to the pyramid tip is missing, and pairs of rotated, similarly looking VB states are obtained instead, as shown in Ref. [37]. Hence, the inclusion of SO interaction is primarily necessary to enable a realistic HH-LH mixing, rather than taking into account strong SO band contributions.

Optical properties

As detailed in Sections 4.4.1 and 6.1, basically all transitions are dipole-allowed in strained pyramidal QDs although they show a large spread in intensity and polarisation. The transition matrix elements from the EP calculation and the present eight-band $\mathbf{k}\cdot\mathbf{p}$ calculation agree within ±10% of the groundstate transition matrix element. The least accurate values belong to the $\tilde{C}1 \rightarrow \tilde{V}2$, $\tilde{C}2 \rightarrow \tilde{V}2$, and $\tilde{C}3 \rightarrow \tilde{V}3$ transitions and are overestimated due to the parallel, rather than orthogonal, alignment of excited CB and VB states shown in Fig. 3.4(a,c). Inclusion of piezoelectricity changes the geometrical alignment, see Fig. 3.4(b,d), and improves the agreement between both calculations.

The in-plane polarization anisotropy (defined as $I_{[1\bar{1}0]}/I_{[110]}$) of the ground-state transition is 0.88 which, compared to the EP value 1.26, is rotated by 90°. The rotation occurs because both $\mathbf{k}\cdot\mathbf{p}$ ground-state wave functions are rotated with respect to the EP calculation. The CB-VB Coulomb matrix elements agree with the EP values within $2.9\pm1.1\,\mathrm{meV}$. Such small deviations may well originate from numerical details or the different approaches for treating the dielectric function.

Section 3.2.1 has shown that both the strain distribution and the strain-independent ASA effect [38] give rise to a C_{2v} symmetry of the bound states.

However, the orientations of the excited CB states differ by 90° for both cases, with consequences on the dipole transition polarizations. Hence, ASA and strain cause opposite effects on the optical properties of InAs/GaAs pyramids.

3.2.2 Inclusion of piezoelectricity

The piezoelectric charging of the QD creates a quadrupole potential shown in Fig. 4.7(d) which is also seen in the CB edge profile in Fig. 3.4(b). It forces the CB (VB) states to prefer the [110] ($[1\bar{1}0]$) direction, see Fig. 3.4(b,d), thus causing a rotation of the first two exited CB states by 90° as compared to the situation without piezoelectricity. Consequently, the piezoelectric effect has a very similar impact on the optical poperties of the model QD as the ASA. The resulting probability densities are shown in Fig. 3.3(c,d) and look virtually identical to those from the EP calculation [38]. In particular, the $V3$ state correctly shows the characteristic LH leakage (”nose”) which is sensitive to the LH/HH band mixing, as discussed above. The projections of the VB states to the HH, LH, and SO bulk bands are shown in Fig. 3.3(e-g) and prove the LH related origin of the ”nose”. The percent values show that the band mixing is hardly influenced by the piezoelectric effect.

The energy levels are shown in Fig. 3.1(c) and exhibit the same rigid shift compared to the EP calculation as observed neglecting piezoelectricity (Fig. 3.1(d)). The CB (VB) levels are 43 ± 6 (39 ± 3) meV lower than in the EP calculation, the ground-state transition energy agrees within 4 meV. $C1$ and $C2$ are split by 9 meV now. The energy difference $C3 - C0$ ($V3 - V0$) agrees with the EP value within 0.3 (3.5) meV. Thus the quantization energies of all considered states agree well with each other. The $\mathbf{k}\cdot\mathbf{p}$ calculated CB-VB transition energies deviate from the EP calculated ones by -4 ± 6 meV only. The transition matrix elements agree with the EP values within ±4% of the ground-state transition matrix element. The ground-state transition in-plane polarization anisotropy is 1.12 and has the same direction as in the EP calculation, like all other anisotropies. The Coulomb matrix elements agree within 2.5 ± 0.5 meV.

Thus, taking into account the piezoelectric effect, a very close agreement between eight-band $\mathbf{k}\cdot\mathbf{p}$ and EP theory is observed. The remaining small differences attributed to the different models and the intrinsic limitations of a given numerical comparison are of no practical relevance.

3.2.3 Smaller quantum dots

A close agreement between the ”generalized” eight-band $\mathbf{k}\cdot\mathbf{p}$ approach and an advanced atomistic model has been proven for the specific case of InAs/GaAs

QDs with 11.3 nm base length. For this size the effects of limited bandstructure representation or the envelope function ansatz are negligible. For InAs/GaAs QDs larger than the model QD (like in most experiments) the accuracy of the $\mathbf{k}\cdot\mathbf{p}$ method will further improve, doubtlessly providing for reliable theoretical predictions.

The accuracy of the eight-band $\mathbf{k}\cdot\mathbf{p}$ model will, however, decrease for smaller dots. The increasing kinetic energy finally brings even the parameter fitting procedure proposed here to its limits. To determine this break-point the above calculations including piezoelectricity are repeated for two smaller InAs/GaAs QDs having pyramid base lengths $b = 9.0\,\mathrm{nm}$ and $b = 6.8\,\mathrm{nm}$, which also were modeled by the EP method [38].

The $\mathbf{k}\cdot\mathbf{p}$ energy levels exhibit the same rigid shift by $\sim 46\,\mathrm{meV}$ compared to the EP calculation as for $b = 11.3\,\mathrm{nm}$. After compensation of this shift, the EP calculated CB (VB) states for $b = 9.0\,\mathrm{nm}$ are reproduced with deviations of $+7\ldots+27\,\mathrm{meV}$ $(-1\ldots+7\,\mathrm{meV})$ and the groundstate transition energy agrees within 8 meV, suggesting that this QD size is still reasonably well described by the eight-band $\mathbf{k}{\cdot}\mathbf{p}$ model. The CB (VB) levels for even smaller dots ($b = 6.8\,\mathrm{nm}$) deviate by $+33\ldots+55\,\mathrm{meV}$ $(-20\ldots-1\,\mathrm{meV})$ from the EP values, and the $\mathbf{k}\cdot\mathbf{p}$ groundstate transition energy is 33 meV larger than the EP value. The transition energies of this dot size are overestimated due to the increasing deviation between the $\mathbf{k}\cdot\mathbf{p}$ and EP bulk bandstructures for $|\mathbf{k}| > 1.8\,\mathrm{nm}^{-1}$, see Fig. 3.2(a). This $\mathbf{k}$-space limit corresponds to a real space size of 3.5 nm which approaches the height of this smallest QD.

Hence, the comparison regarding the InAs/GaAs material system shows a reasonable agreement between eight-band $\mathbf{k}{\cdot}\mathbf{p}$ and EP calculations for $b \geq 9\,\mathrm{nm}$. The contributions from "large-$|\mathbf{k}|$ states" are apparently negligible in this size range. The valid QD size range for eight-band $\mathbf{k}{\cdot}\mathbf{p}$ calculations depends on the material system, of course. However, the predictive power can be checked in advance by determining the minimal feasible structural size from the maximal feasible $|\mathbf{k}|$ value, using bandstructure diagrams like in Fig. 3.2.

3.3 Resumé

In order to improve the accuracy of the eight-band $\mathbf{k}\cdot\mathbf{p}$ model in electronic structure calculations of strongly confining and inhomogeneously strained QDs, it is suggested to determine all $\mathbf{k}\cdot\mathbf{p}$ parameters from a fit to prevailing data on the bulk bandstructures of all constituent materials in the heterostructure. The goal of the fit procedure is the reproduction of a given bulk bandstructure by the $\mathbf{k}\cdot\mathbf{p}$ model within an as large as possible region of the BZ.

The size of this **k**-space region should correspond to the minimal structural size in the real space. Thus the QDs must not be too small for typical semiconductors. In the case of InAs/GaAs QDs high accuracy is achieved for pyramid base lengths $b \geq 9\,\mathrm{nm}$, matching well the experimentally known size range of self-organized InAs QDs. Below this size, transition energies become increasingly overestimated due to an increasing deviation between the true and the **k**·**p** bandstructure. Second, linear deformation potentials used in the Pikus-Bir type Hamiltonian should be adjusted as to fit the average large-strain situation in the QDs, otherwise systematic deviations due to non-linearities in the true system may occur. Alternatively, the eight-band **k**·**p** model is easily extended to include hydrostatic band gap non-linearities.

The enhancement of accuracy achieved by the proposed fit procedure was demonstrated by exemplaric calculations for InAs QDs. The eight-band **kp** model including the piezoelectric effect predicts the same electronic and optical properties for pyramidal InAs/GaAs quantum dots as the currently most advanced microscopic theory. The quantitative agreement between both approaches extends even to fine details of band mixing effects in wave functions and yields an overall accuracy of $\sim 10\,\mathrm{meV}$ for exciton transition energies. For practical electronic structure calculations, in particular for comparison with experimental results, the eight-band **k**·**p** method is fully valid. Its strength is the flexibility in treating realistic, large structures and (qua)ternary alloys like In(Ga)As/Ga(Al,P)As, capturing the important effects of strain, piezoelectricity, and band mixing. This renders the model ideally suited for field studies of the correlation of structural and optical properties of strained, low-symmetry QD heterostructures.

Part II

Electronic Structure of Quantum Dots and Wires

Chapter 4

Electronic Structure of Quantum Dots

Three regimes of confinement may be distinguished by comparing the effective radius R of the QD with the Bohr radii[1] of electrons (a_B^e) and holes (a_B^h) in the respective bulk material [1, 85, 86], where typically $a_B^e > a_B^h$: $R > a_B^e$ defines weak confinement and leads to a quantization of the center-of-mass motion of excitons while the exciton binding energy E_X is mainly due to Coulomb interaction. $a_B^e \geq R \geq a_B^h$ defines the intermediate confinement regime where mainly the electron is quantized, but not the hole. In this case quantum confinement and Coulomb interaction have comparable influence on E_X. Finally, $a_B^h > R$ defines strong confinement where both carrier types become quantized and E_X is strongly increased by the structural confinement.

Already from a substantially simplified viewpoint employing effective-mass theory with parabolic bands it becomes clear that three-dimensional (3D) strong confinement lifts any $\mathbf{k}$ conservation in the bound states of charge carriers: The Hamiltonian $\hat{H} = \hbar^2\mathbf{k}^2/(2m) + V(\mathbf{r})$ does not commute with $\mathbf{k}$ in any spatial direction. Hence, neither $\mathbf{k}$, nor the momentum $\mathbf{p}$ yield useful observables in such systems. In absence of magnetic fields bound states in QDs are degenerate by time-reversal symmetry [87, 88].

According to single-band effective-mass theory the single-electron and -hole energies in a rectangular QD (edge lengths $L_x \neq L_y \neq L_z$) with an infinitely high potential barrier V (the extreme case of strong confinement) are given by

$$E^{e,h}_{\alpha\beta\gamma} = E^{c,v} \pm \frac{\pi^2\hbar^2}{2m^{e,h}}\left(\frac{\alpha^2}{L_x^2} + \frac{\beta^2}{L_y^2} + \frac{\gamma^2}{L_z^2}\right) \quad , \tag{4.1}$$

---

[1] The Bohr radius serves as an elementary length unit and is defined here as $a_B = \hbar^2/(me^2)$ [84] where m is the Γ-point effective mass of the CB or VB, respectively.

[89] where "+" ("−") holds for electrons (holes), $\alpha, \beta, \gamma = 1, 2, \ldots$ are integer quantum numbers, and E^c and E^v denote the band edges of the bulk material forming the QD. If the QD is spherical with radius R then

$$E_{nl}^{e,h} = E^{c,v} \pm \frac{\hbar^2}{2m^{e,h}} \left(\frac{\chi_{nl}}{R} \right)^2 \tag{4.2}$$

[4] where $n = 1, 2, \ldots$ is the radial and $l = 0, 1, \ldots$ the angular quantum number, and χ_{nl} is the n-th zero of the spherical Bessel function j_l. An energy level scheme according to Eq. (4.2) is shown in Fig. 4.1(a). For both dot shapes, the effective-mass theory predicts simple selection rules for possible dipole transitions between electrons and holes ($\Delta\alpha = \Delta\beta = \Delta\gamma = 0$ and $\Delta n = \Delta l = 0$, respectively) as well as for transitions between multi-exciton states [89].

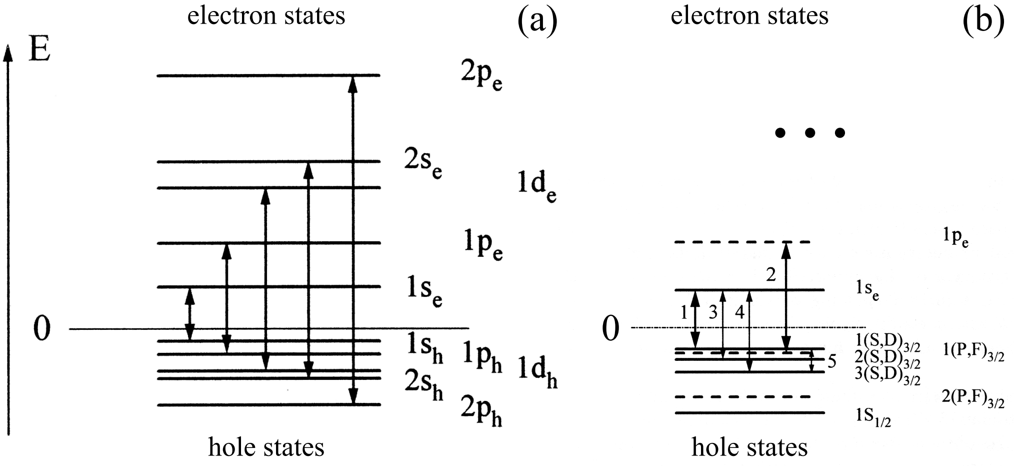


Figure 4.1:

Energy levels and allowed single-particle dipole transitions in a spherical QD with infinite barrier, according to effective-mass theory (a) using parabolic bands and yielding Eq. (4.2), and (b) including VB mixing (taken from Ref. [4]).

This picture is, however, too simplified to apply to real 0D semiconductor structures. Apart from essentially different geometries encountered in lithographically defined or self-organized epitaxial QDs, realistic energy level schemes require modifications taking into account band-mixing effects, finite barrier heights, strain effects and, if applicable, piezoelectricity, as already done in Chapter 3.

According to Eq. (4.2) there is orbital angular momentum conservation in the spherical QD. This emerges from the *mesoscopic* central symmetry of V, which is artificially preserved on the *microscopic* (i. e., atomic) scale by assuming parabolic bulk bands in effective-mass theory. In actual fact the valence bands are not even isotropic, let alone parabolic (see Fig. 2.3 and Fig. 3.2), so that the mesoscopic and microscopic symmetries of the confinement do not match. This causes a mixing of, at least, the HH and LH bands in the 0D hole wavefunctions, with the consequence of fundamentally altered wavefunction symmetries and 0D hole energies. The remaining conservation values for spherical QDs are the parity and the *total* angular momentum, the latter of which merges the mesoscopic orbital angular momentum and the microscopic angular momentum of the Bloch functions [4]. Fig. 4.1(b) shows an energy level scheme of a spherical QD taking into account VB mixing. The modified hole wavefunction symmetries also lead to different dipole selection rules as compared to Fig. 4.1(a): Formerly forbidden transitions with $\Delta n \neq 0$ acquire significant oscillator strength due to the VB mixing. In materials with small SO energies Δ_0, like InP or CdS, the mixing will also involve the SO band [90]. In narrow-gap materials like InAs even the 0D electron levels are affected by band mixing, as demonstrated in Section 3.2.1.

In buried QDs (being surrounded by another semiconductor material) the confining barrier V is of finite height as defined by the band offsets. Hence there is a finite number of bound states strongly depending on the QD size, as well as on the geometrical shape and chemical composition. As the QDs become smaller the energy separations between the levels increase (quantum size effect), and below a certain size no bound state will exist at all.

In QDs formed in a lattice-mismatched material system like InAs/GaAs, neither the barrier heights along the heterointerface, nor the confinement potential values $V(\mathbf{r})$ inside the QD are constant, due to pseudomorphic strain: The spatially varying strain field modulates the band edges and thus their discontinuities at each position $\mathbf{r}$ along the heterointerface. An additional contribution to the confinement potentials may arise from a piezoelectric charging of the QD heterostructure, scaling linearly with the QD size (to first order). Hence the confinement itself is size dependent. In such QDs the two conservation values still present in Fig. 4.1(b) do not exist anymore, and there are no good quantum numbers which persist during a change of the QD size. Finite-barrier effects and band mixing lead to a lifting of dipole selection rules, to pronounced optical polarization properties, and to an enhancement of mid-infrared optical transitions between the 0D electrons or holes, respectively.

Few years ago the first experimental evidence using temperature dependent cathodoluminescence experiments was obtained for the existence of 0D electronic states in self-organized InAs/GaAs QDs [91] which had been fabricated by Stranski-Krastanow growth during molecular beam epitaxy. Since then this type of dots has emerged as one of the most extensively studied QD systems [1]. Today, a large number of experimental results is available addressing the epitaxy [92], as well as the structural, optical [93], and transport [94,95] properties of such dots. QD lasers operate at room temperature both cw and pulsed up to $\approx 10\,\mathrm{GHz}$ [17,18,96].

The size and shape of Stranski-Krastanow grown InAs QDs on GaAs (001) vary dependent on the epitaxial method and, particularly, on the growth conditions used: Pronounced pyramidal shape or a less pointed dome shape, side facets oriented along $\{101\}$ [97], $\{105\}$ [98], $\{113\}$ [99], $\{114\}$ [100,101], $\{136\}$ [102], rather flat disk-like structures [103], and different sizes of QDs [12] have been observed.

Accordingly, low temperature spectra of a few or single QDs exhibit rich excitonic structures, but differ according to the growth conditions [12,104–109]. In larger ensembles of QDs inhomogeneous line broadening of presently $\geq 18\,\mathrm{meV}$ [12] due to fluctuations of size, shape and chemical composition disguises part of this spectral structure. Nevertheless, transitions due to excited exciton states are clearly visible in calorimetric absorption [91], photoluminescence (PL) [110], and photoluminescence excitation (PLE) spectra [12,111–113], indicating the existence of several excited levels for both electrons and holes in self-organized InAs/GaAs dots. Unambiguous assignment of the transitions is still pending, although progress has been made, advancing well beyond interpretations based on two-dimensional harmonic oscillator models (as detailed in Chapter 6.3). In order to obtain meaningful interpretations, advanced numerical bandstructure calculations need to be correlated with precise information about the structural properties of the QDs [12].

Large effort has been spent since 1994 [114,115] at the electronic structure calculation for this QD system. The dot shape was often assumed to be a pyramid with $\{101\}$ side facets, like in this chapter, in agreement with results in Ref. [97]. The energy levels have been calculated using conventional effective mass models [24,114], perturbational effective mass approaches [20,21,23], eight-band $\mathbf{k}\cdot\mathbf{p}$ theory [25–32], tight-binding [35], and EP theory [36–39]. Thereby, varying predictions for both the number and actual energies of the levels were obtained, being mainly due to varying assumptions about the bulk properties of the constituent materials: As demonstrated in Chapter 3, eight-band $\mathbf{k}\cdot\mathbf{p}$ and EP theory yield quantitatively equivalent predictions if the same bulk properties are assumed in both calculations, and differences between various $\mathbf{k}\cdot\mathbf{p}$ calculations reported are primarily due do different material parameters used [31].

In the following, the size dependence of the electronic properties of capped, single pyramidal InAs QDs with $\{101\}$ facets on GaAs (001) is investigated for the QD base length b range $10\,\mathrm{nm} \leq b \leq 20\,\mathrm{nm}$ using the eight-band $\mathbf{k}\cdot\mathbf{p}$ model. The influence of selected material parameter variations on the single-particle energies is discussed. Also the influence of a truncation of the pyramid on the electronic properties was studied, and published in [116]. A similar study concerning truncated InAs pyramids was reported in Ref. [32]. A size-dependence study of pyramidal InAs dots, based on EP calculations, was reported in Ref. [38].

Since the impact of strain on the carrier confinement is comparable to that of the band offset due to the variation of the chemical composition at the heterojunctions, the wave functions and energies are sensitive to the underlying strain distribution (see also Section 3.1.1). Therefore, the influence of the chosen strain model on the energy levels is investigated next: The continuum mechanical (CM) model using the (experimentally) known elastic moduli C_{11}, C_{12}, C_{44} is compared to a linearized valence force field (VFF) model, and to a CM calculation employing an incorrect value of C_{44} implicitly assumed in the VFF model.

4.1 Calculation of strain

The software for strain calculations according to the CM or VFF model was developed and written by Marius Grundmann.

4.1.1 Continuum mechanical model

The total strain energy in the CM model is given by [57]

$$U_{\mathrm{CM}} = \frac{1}{2} \sum_{i,j,k,l} C_{ijkl} \epsilon_{ij} \epsilon_{kl} \qquad . \tag{4.3}$$

For a given structure U_{CM} is minimized, using finite differences for the strains $\epsilon_{ij} = \partial u_i / \partial x_j$, where $\mathbf{u}$ is the displacement vector field. The elastic moduli C_{ijkl} are represented by the parameters C_{11}, C_{12}, and C_{44} for cubic crystals. The strain distribution in capped InAs pyramids with $b = 13.6\,\mathrm{nm}$ on a thin InAs WL in GaAs, shown in Fig. 4.2 and Fig. 4.3, is calculated using $\sim 2.6 \cdot 10^6$ voxels and a conjugate-gradient method for the total energy minimization. At interfaces, the condition of a continuous stress tensor yields the proper boundary conditions. In order to avoid oscillatory solutions arising when symmetric difference quotients are used, the energies from the eight possible combinations of forward and backward differences in the three directions are averaged.

4.1.2 Valence force field model

Sufficiently small QDs can also be directly modeled with the VFF model [117–120]. The elastic energy of each atom is written in terms of the positions of its nearest neighbor atoms and then added up for all atoms. In the version of

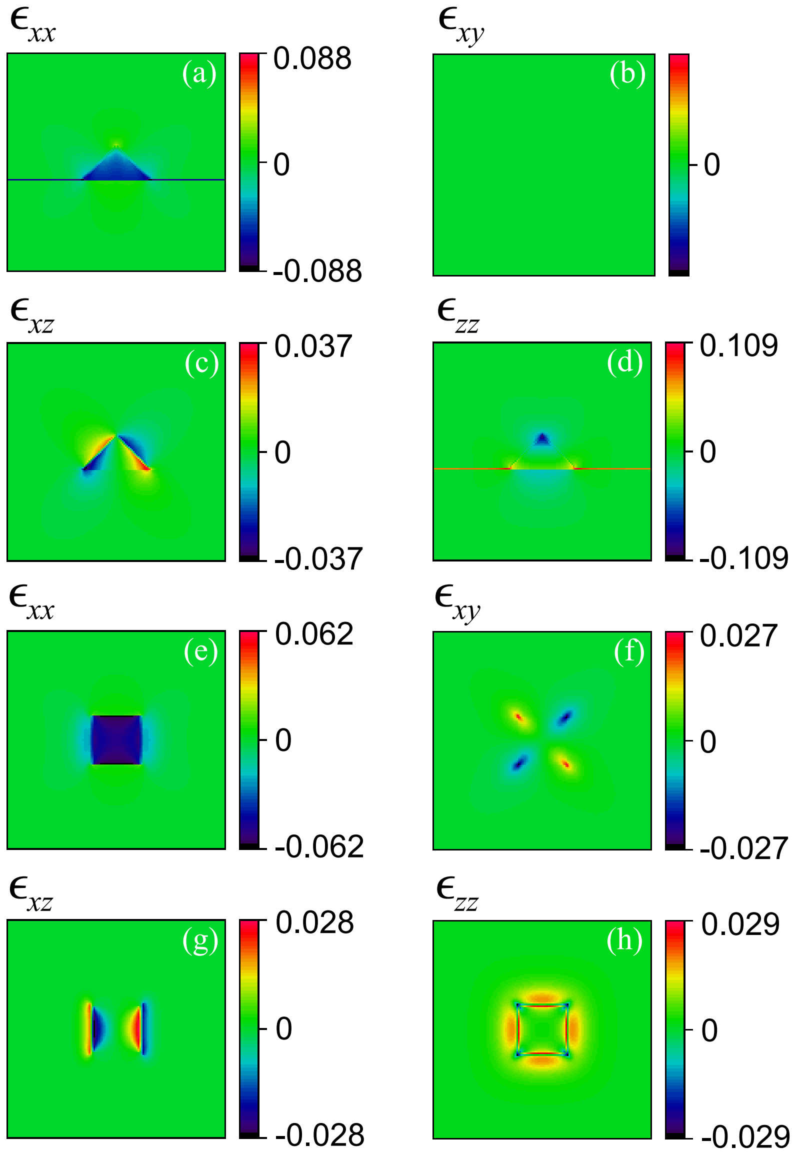


Figure 4.2:

Strain distribution in and around an InAs pyramid ($b = 13.6\,\text{nm}$) bounded by $\{101\}$ facets on a 1 ML thick InAs WL within a GaAs matrix, calculated according to the CM model. The strain tensor components ϵ_{xx} (a,e), ϵ_{xy} (b,f), ϵ_{xz} (c,g), and ϵ_{zz} (d,h) are shown in the (100) cross section planes through the pyramid tip (a-d) and in the (001) cross section plane a few MLs above of the WL (e-h).

Keating [117], the elastic energy U_{VFF} of the crystal is written as a sum over all atoms i and given as

$$U_{\text{VFF}} = U_\alpha + U_\beta \quad , \tag{4.4}$$

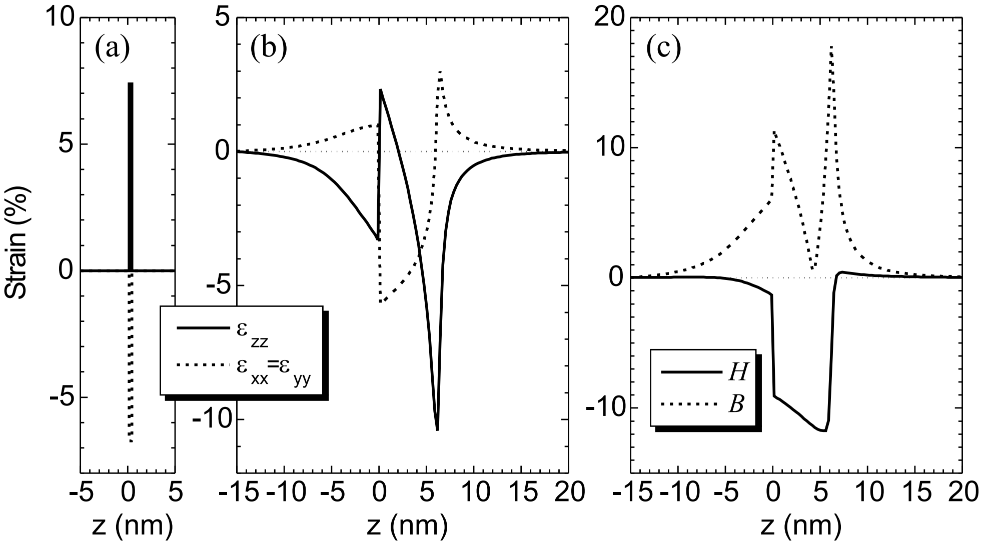


Figure 4.3:

Strain distribution in and around an InAs pyramid ($b = 13.6\,\text{nm}$) bounded by $\{101\}$ facets on a 1 ML thick InAs WL within a GaAs matrix, calculated according to the CM model and plotted along [001] linescans: (a) through the WL far away from the dot, (b) and (c) through the pyramid tip. (a) and (b) show the diagonal elements of the strain tensor and (c) depicts the hydrostatic (H) and biaxial (B) strain components: $H = \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}$, $B^2 = (\epsilon_{xx} - \epsilon_{yy})^2 + (\epsilon_{yy} - \epsilon_{zz})^2 + (\epsilon_{zz} - \epsilon_{xx})^2$.

$$U_\alpha = \frac{1}{4}\sum_i \left[\frac{\alpha_{ij}}{4}\sum_j \frac{(\mathbf{r}_{ij}\cdot\mathbf{r}_{ij} - 3d_{ij}^2)^2}{d_{ij}^2}\right] \quad , \tag{4.5}$$

$$U_\beta = \frac{1}{4}\sum_i \left[\sum_j \sum_{k\neq j} \frac{\beta_{ijk}(\mathbf{r}_{ij}\cdot\mathbf{r}_{ik} + 3d_{ij}d_{ik})^2}{2d_{ij}d_{ik}}\right] \quad , \tag{4.6}$$

where the sums over j and k run over the four tetrahedrally coordinated nearest neighbor atoms, $\mathbf{r}_{ij}$ denotes the vector from the ith atom towards its jth neighbor, and $4d_{ij}$ is the lattice constant of the binary (or elementary) $i-j$ constituent. U_α in Eq. (4.4) is non-zero when the bond *length* is changed from the strain-free state and is thus called "bond-stretching" interaction. U_β in Eq. (4.4) is non-zero when the *angle* between bonds is altered and is thus called "bond-bending" interaction.

By comparison with the cubic strain tensor it follows that α and β can be expressed in terms of C_{11} and C_{12} [117]:

$$\alpha = (C_{11} + 3C_{12})d \quad , \tag{4.7}$$

$$\beta = (C_{11} - C_{12})d \quad . \tag{4.8}$$

	Si	Ge	GaAs	InAs	InP
κ	0.99	1.07	1.13	1.22	1.20

Table 4.1: Ratio κ according to Eq. (4.10) for different semiconductors.

In a ternary compound or across a heterointerface the β parameter is geometrically averaged if the atoms j and k are not identical: $\beta_{ijk} = \sqrt{\beta_{ij}\beta_{ik}}$ [121]. Since this VFF model works with two parameters only, C_{44} is no independent elastic modulus anymore but fixed to the value

$$C_{44} = \frac{\alpha\beta}{(\alpha+\beta)d} \quad , \tag{4.9}$$

$$\text{i. e.} \qquad \kappa = \frac{2C_{44}(C_{11}+C_{12})}{(C_{11}-C_{12})(C_{11}+3C_{12})} \equiv 1 \quad . \tag{4.10}$$

This relation is fulfilled for silicon. For a number of other important semiconductors there are deviations up to 22% (InAs), see table 4.1.

The potential of Eq. (4.4) is not harmonic, and this unharmonic Keating model has been used in Chapter 3, motivated by Ref. [37–39,122]. However, as remarked by Kane [120], anharmonic effects due to the higher order terms have not been shown to be satisfactorily treated by this Keating model. In particular, the third-order elastic moduli C_{ijk} [123, 124] do not enter the theory. Therefore, the linearized version of U_α and U_β in Eq. (4.4), as proposed in Ref. [120], is used in this work.[2]

In addition, second-nearest neighbor bond stretching, contiguous bond bending, and the so called MSBN interaction [126, 127] can be included to the VFF model. The eight parameters of Solbrigs model [126, 127] are determined from fits to experimental phonon dispersion curves. It has been applied to InAs/GaAs pyramids [24] but is not pursued here in order to avoid its excessive numerical expense. Instead, two more easily tractable models are compared: CM and the linearized Keating model.

The differences between the strain distributions in a pyramid calculated within the CM and the (linearized) VFF models are shown in Fig. 4.4. Two different continuum mechanical models are compared, first CM with the correct value of C_{44}, and second CM[C_{44}^{VFF}] where C_{44} takes the incorrect value implied by the two-parameter Keating model according to Eq. (4.9). The main differences between the VFF and CM models are:

---

[2] Most recently, after the submission of this thesis, a generalized VFF model taking also into account C_{44} has been proposed [125].

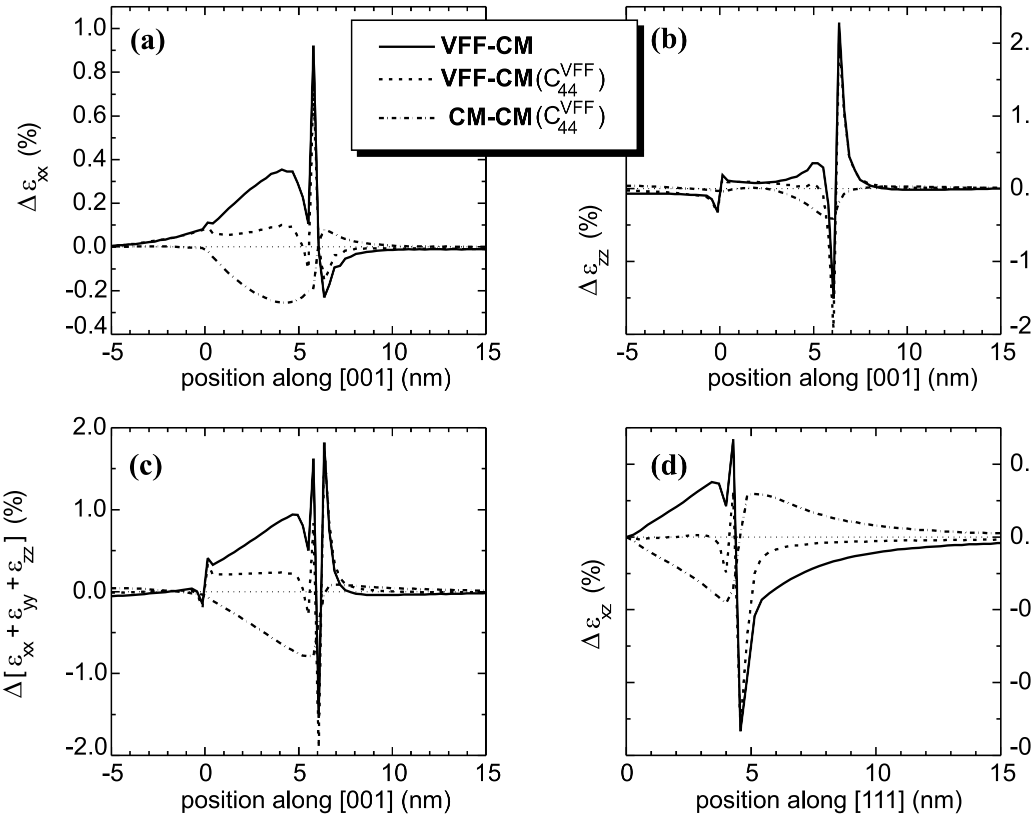


Figure 4.4:

Strain distributions in an InAs pyramid ($b = 13.6\,\mathrm{nm}$) on a 1 ML thick WL. The differences between the (linearized) VFF model, the CM model, and CM model with the value for C_{44} according to Eq. (4.9) used in VFF (CM[C_{44}^{VFF}]) are compared. (a) ϵ_{xx}, (b) ϵ_{zz}, (c) hydrostatic strain along [001] through the center of the pyramid, (d) shear component ϵ_{xz} along [101], originating at the bottom center of the pyramid. Solid (dashed) lines represent the difference between the VFF and CM (CM[C_{44}^{VFF}]) models. Dash-dotted lines show the difference between the CM and CM[C_{44}^{VFF}] models.

- At the tip of the pyramid the diagonal strain components ϵ_{ii} analytically diverge in continuum theory while the VFF model yields a finite value. A numerical solution of the CM model using finite differences with a voxel size of one atom yields a similar value, however.
- At interfaces the VFF and CM models differ on the atoms directly adjacent to the interface, as shown in Fig. 4.5.
- In the volume, differences between VFF and CM (see Fig. 4.5) are mainly due to the incorrect value of C_{44} in VFF. They almost disappear if VFF is compared to CM[C_{44}^{VFF}] which uses the incorrect C_{44} value according to Eq. (4.9).

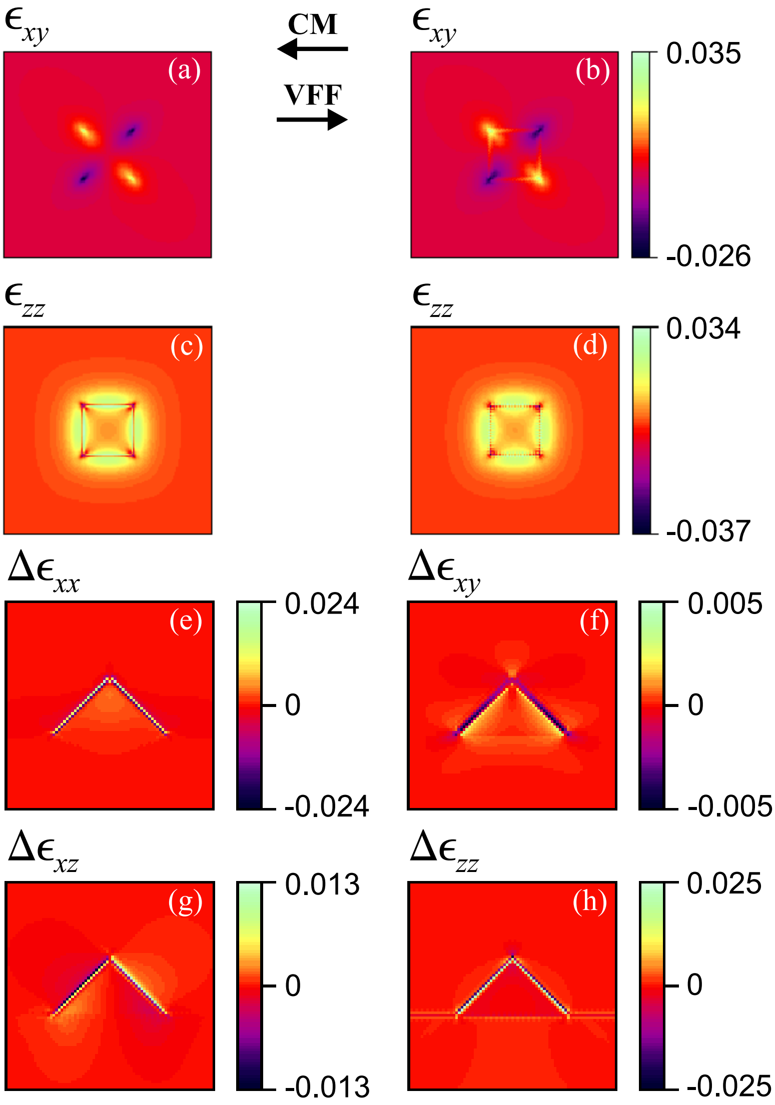


Figure 4.5:

Comparison between the CM and VFF strain models. The strain tensor components ϵ_{xy} and ϵ_{zz} as calculated by the CM (a,c) and VFF (b,d) models, respectively, are plotted along a (001) cross section plane somewhat above the WL. The numerical *differences* between the two models, $\Delta\epsilon = \epsilon_{\mathrm{VFF}} - \epsilon_{\mathrm{CM}}$ are shown for the strain tensor components ϵ_{xx} (e), ϵ_{xy} (f), ϵ_{xz} (g), and ϵ_{zz} (h).

- The symmetry of the CM strain tensor in the (001) plane is C_{4v}. The tetrahedral configuration of atoms in the VFF model leads to a C_{2v} symmetry, i. e. strain components are different along the [110] and [1$\bar{1}$0] directions. This is visualized in Fig. 4.6, where the difference of the strain component

ϵ_{zz} along the [110] and [1$\bar{1}$0] directions is plotted. In the CM model (for any value for C_{44}) this difference is *a priori* zero.

The influence of these differences on the energy levels and wave functions of charge carriers in the QD is discussed below.

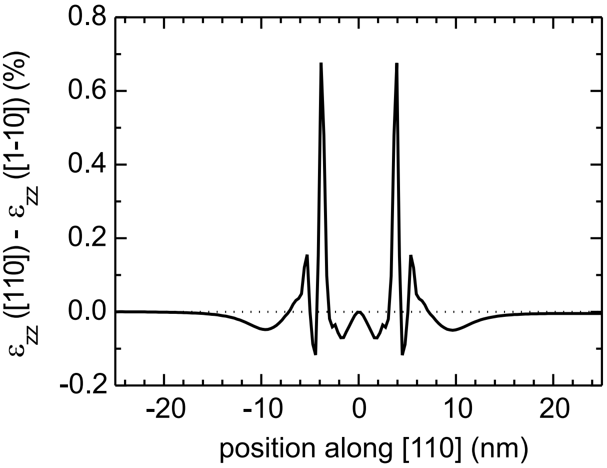


Figure 4.6:

Difference of ϵ_{zz} along the [110] and [1$\bar{1}$0] directions along a [110] linescan through the center of an InAs pyramid ($b = 13.6\,\mathrm{nm}$) on a 1 ML thick WL, calculated using the VFF model.

4.2 Piezoelectricity

Regardless of the chosen strain model, the symmetry of the carrier confinement is only C_{2v} due to piezoelectric [74–78] charging of the QD [114]. From the strain tensor (ϵ_{ij}) and the piezoelectric modulus e_{14} the piezoelectric charge density

$$\rho_p(\mathbf{r}) = \mathrm{div}\,[e_{14}(\mathbf{r}) \cdot \{\epsilon_{yz} + \epsilon_{zy}\,,\, \epsilon_{xz} + \epsilon_{zx}\,,\, \epsilon_{xy} + \epsilon_{yx}\}(\mathbf{r})] \tag{4.11}$$

is calculated. Eq. (4.11) introduces C_{2v} symmetry to the Hamiltonian Eq. (4.14) even when (ϵ_{ij}) possesses C_{4v} symmetry: The polarization charges mainly form dipoles along the edges of the pyramid and the polarities of the dipoles alter between inside and outside the QD as well as between neighbouring edges, see Fig. 4.7(a). At edges along [1 $\bar{1}$ 1] and [1 $\bar{1}$ $\bar{1}$] the negative charges are located inside the QD. This implies that the [1$\bar{1}$0] direction will be the favourable direction for the extension of hole wave-functions, while electrons will prefer to expand along [110].

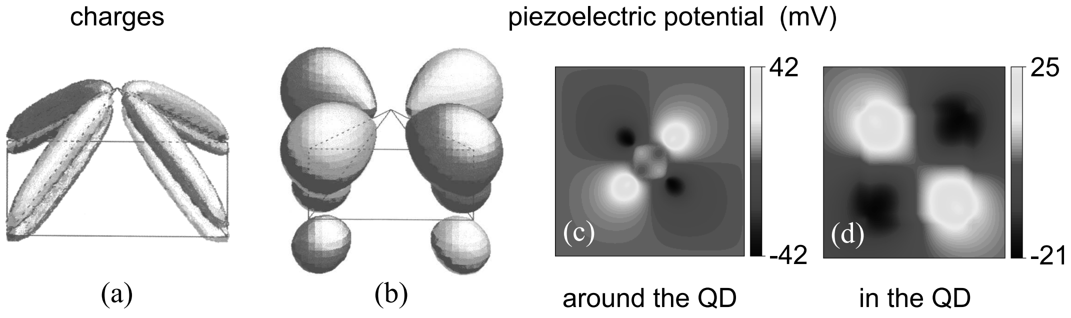


Figure 4.7:

Piezoelectric charges (a) and their Coulomb potential (b-d) in a pyramidal InAs QD with $b = 12\,\mathrm{nm}$. (a): Isosurfaces corresponding to volume charges of $\pm 0.3\,e\,\mathrm{nm}^{-3}$ [114], light gray denotes positive charge and dark gray negative charge. (b): Associated Coulomb potential isosurfaces [114], light gray denotes $+30\,\mathrm{meV}$ and dark gray $-30\,\mathrm{meV}$. (c,d): Cross section through the Coulomb potential along a (001) plane somewhat above the WL in two different magnifications, (d) is a zoom into (c). The InAs/GaAs heterointerface is visible by means of the image charge effect.

The associated Coulomb potential $V_p(\mathbf{r})$ is obtained from the Poisson equation Eq. (4.12),

$$\begin{aligned} & \rho_p(\mathbf{r}) &=& \; \epsilon_0 \nabla \left[\epsilon_s(\mathbf{r}) \nabla V_p(\mathbf{r})\right] & \quad (4.12) \\ \Leftrightarrow \quad & \Delta V_p(\mathbf{r}) &=& \; \frac{\rho_p(\mathbf{r})}{\epsilon_0 \epsilon_s(\mathbf{r})} - \frac{1}{\epsilon_s(\mathbf{r})} \nabla V_p(\mathbf{r}) \cdot \nabla \epsilon_s(\mathbf{r}) \qquad , & \quad (4.13) \end{aligned}$$

where $\epsilon_s(\mathbf{r})$ is the static dielectric constant of the respective material at position $\mathbf{r}$ [128, 129]. The first term on the right-hand side of Eq. (4.13) refers to the true three-dimensional charge density while the second is the contribution of polarization surface charge densities due to discontinuous dielectric constants at hetero-interfaces (the singularity of $\nabla \epsilon_s$ is integrable), usually discussed as image charge effect image charges [130–132]. Numerically, Eq. (4.12) needs to be solved on a substantially larger space region than the bandstructure equation Eq. (4.14), and using Dirichlet boundary conditions. Otherwise a too small spatial modulation of $V_p(\mathbf{r})$ may be obtained.

The piezoelectric potential $V_p(\mathbf{r})$ is shown in Fig. 4.7(b-d) and scales essentially proportionally to the QD size b. In the studied dot size range the piezoelectric influence on the bound particle states dominates over that of immediate C_{2v} symmetry effects from the VFF strain distribution, as also discussed in Chapter 3.

4.3 Bound single-particle states

The energy levels and wave functions of bound electron and hole states are calculated using the 8-band **k·p** model outlined and validated in Part I. The spatial representation of the Hamiltonian H yields a system

$$H\left(x,y,z,\partial_x,\partial_y,\partial_z,\partial_{xx},\partial_{xy},\partial_{xz},\partial_{yy},\partial_{yz},\partial_{zz}\right)\Psi(x,y,z) \;=\; E\,\Psi(x,y,z) \tag{4.14}$$

of eight coupled partial differential equations for the eight complex envelope functions

$$\Psi = \left(\psi_{s\Uparrow},\psi_{x\Uparrow},\psi_{y\Uparrow},\psi_{z\Uparrow},\psi_{s\Downarrow},\psi_{x\Downarrow},\psi_{y\Downarrow},\psi_{z\Downarrow}\right) \qquad . \tag{4.15}$$

Those eigensolutions (E_n, Ψ_n) of Eq. (4.14) that vanish on the boundary ∂G of the calculation region G, i. e. the integrable ones, are the bound states of charge carriers in the QD. The Hamiltonian is discretized using the finite-differences method described in Chapter 8 and the discrete eigenproblem is solved by an algorithm described in Section 9.3.

The **k·p** material parameters used in this chapter are summarized in Table 4.2. These are conventional literature values which not have been derived according to the procedure proposed in Chapter 3. The reason for neglecting that procedure here is that the predictions obtained in the following, using the parameters from Table 4.2, match significantly better recent experimental results [12] than the EP calculations [36–39] used for comparison in Chapter 3, as discussed in Chapter 6.3. The proposal of Chapter 3 is more methodological, proving the sufficiency of the eight-band **k · p** *model* for the current purpose, while the actually most plausible bulk bandstructures for InAs and GaAs are difficult to determine in practice.

The electron and hole level energies and wavefunctions were calculated for four different QD sizes (base lengths b: 10.2 nm, 13.6 nm, 17.0 nm, 20.4 nm) and using two strain models (CM and VFF). The InAs/GaAs band alignment (see E_v' in Table 4.2) is assumed as suggested by Fig. 2.3, i. e., the offset in the unstrained bulk VB edge is 186 meV (referred to as lineup "A", $\Delta E_c/\Delta E_g = 0.83$). The assumed temperature is 6.5 K.

4.3.1 Results for strain according to the CM model

Fig. 4.8 shows the energy levels for the strain distributions obtained by the CM model. The WL ground-state levels shown are taken from Ref. [114], in agreement with experimentally observed transition energies, and are assumed to be independent from the QD size.

	Unit	Value for $\mathrm{In}_c\mathrm{Ga}_{1-c}\mathrm{As}$	Ref.
a	Å	$5.6503 + 0.4050c$	[69]
E_0	meV	$1518 - 1580c + 475c^2$	[69]
E'_v	meV	$-6920 + 231c - 58c^2$	[40,69]
Δ_0	meV	$340 - 93c + 133c^2$	[45]
E_p	meV	$(1.238 - 0.2095c)\frac{1-m_e}{m_e}\frac{3E_0(E_0+\Delta_0)}{3E_0+2\Delta_0}$	[53,133]
m_e	m_0	$0.0667 - 0.0419c - 0.00254c^2$	[62]
γ_1		$1/[(1-c)/7.10 + c/19.7]$	[45,61]
γ_2		$1/[(1-c)/2.02 + c/8.4]$	[45,61]
γ_3		$1/[(1-c)/2.91 + c/9.3]$	[45,61]
B	$\mathrm{meV\,nm}^2$	0	[33,40]
b'	meV	0	[33,40]
a_c	meV	$-8013 + 2933c$	[62]
a_g	meV	$-8233 + 2153c$	[62]
b_v	meV	$-1824 + 24c$	[45]
d_v	meV	$-5062 + 1462c$	[45]
C_{11}	GPa	$118.8 - 35.5c$	[45,61]
C_{12}	GPa	$53.8 - 8.5c$	[45,61]
C_{44}	GPa	$59.4 - 19.8c$	[45,61]
ϵ_s		$13.18 + 1.42c$	[62]
e_{14}	$\mathrm{C\,m}^{-2}$	$0.160 - 0.115c$	[62]

Table 4.2:

Indium gallium arsenide material parameters for 6.5 K used for the QD calculations in this part of the work.

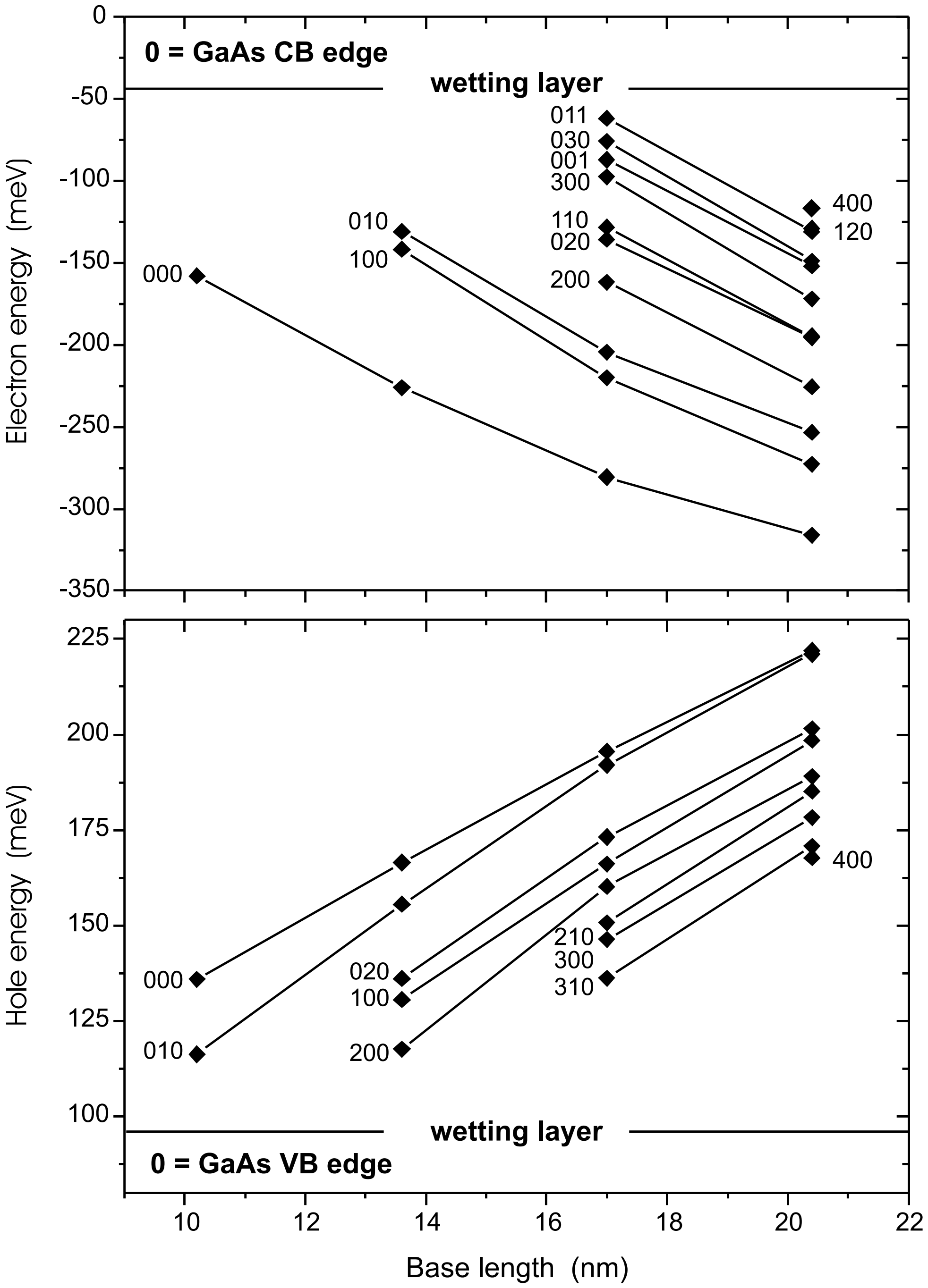


Figure 4.8:

Electron and hole energies in pyramidal InAs/GaAs QDs for different sizes, calculated using the CM strain model. The lines connect levels having the same wave function symmetry, labeled by the appropriate "quantum numbers" (see the text on p. 50 for their explanation).

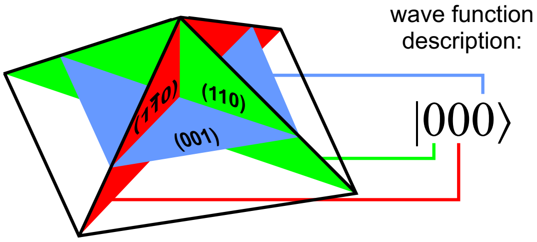
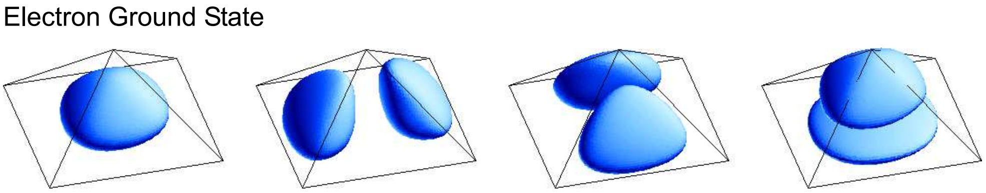
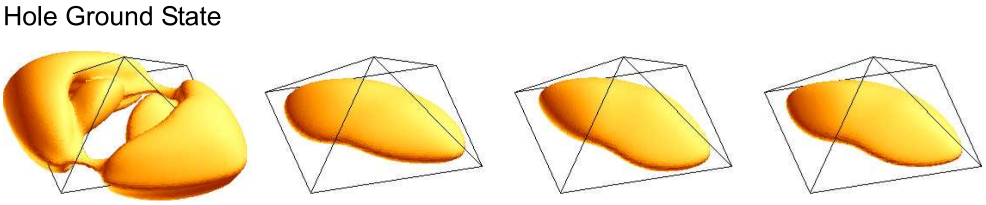


Figure 4.9:

Tangential planes of the probability-density nodal surfaces used to label bound states.



s part: 0.89 p_x part: 0.03 p_y part: 0.03 p_z part: 0.05



s part: 0.01 p_x part: 0.46 p_y part: 0.46 p_z part: 0.07

Figure 4.10:

Decomposition of the ground-state wave functions into their constituing single envelopes (b = 13.6 nm). The numbers denote the integral probability of the respective envelopes after summation over both spins. The electron ground state is mainly represented by s-type Bloch functions while the hole ground state is almost completely described by p-type Bloch functions.

The "quantum numbers" are based on a depictive terminology to graphically describe the shape of the probability density, and refer to the tangential planes of the probability-density nodal surfaces in the order [110], $[1\bar{1}0]$, [001] (see Fig. 4.9). These are, however, no true quantum numbers: As a consequence of band mixing, nodal surfaces exist at most for single envelopes and usually have different orientations for the p-type envelopes $|x\rangle$, $|y\rangle$, and $|z\rangle$ (see Fig. 4.10). The probability density distributions obtained by (weighted) summation over all envelope contributions have no nodal surface, hence the states possess no well-defined parity, see also the discussion in Section 3.2.1.

Due to the piezoelectric field shown in Fig. 4.7(d) the main directions of extension of the wave functions are [110] for electrons, see Fig. 4.11(a) (as well as Fig. 3.3(c) and Fig. 3.4(b)), and $[1\bar{1}0]$ for holes, see Fig. 4.11(b,c) (as well as Fig. 3.3(d) and Fig. 3.4(d)), while $[ij1]$ only occurs in high energy states in large QDs, see the electron state $|001\rangle$ in Fig. 4.11(a).

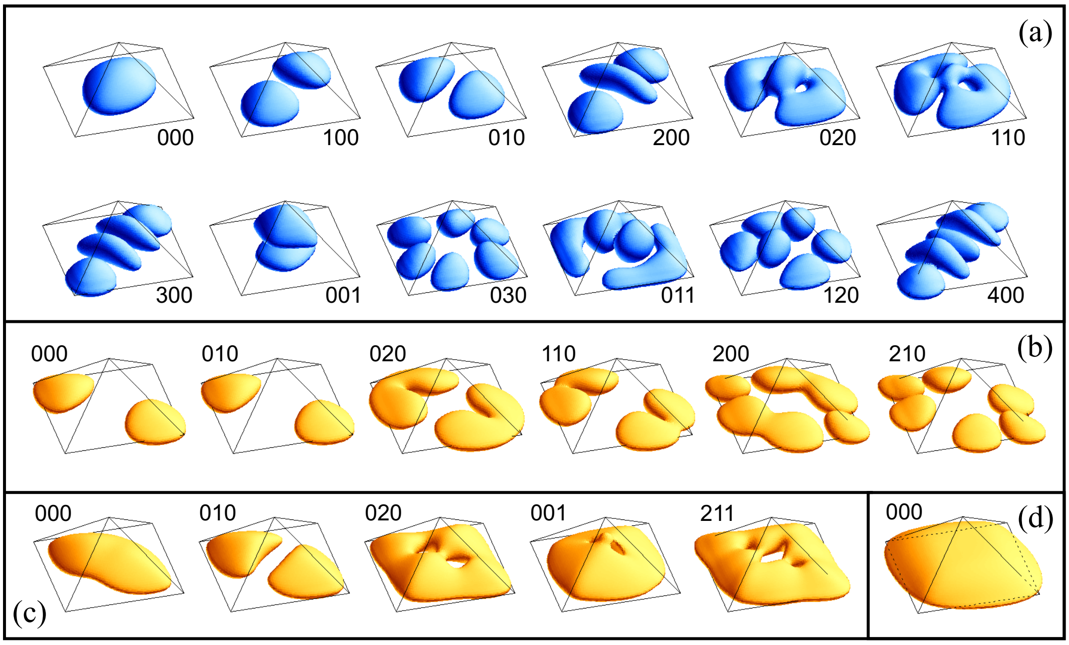


Figure 4.11:

Probability density isosurfaces ($p = 65\%$) of (a) the electron and (b) the hole states for $b = 20.4\,\mathrm{nm}$, the strain is calculated using the CM model. (c) Hole states for $b = 13.6\,\mathrm{nm}$ and strain calculated using the VFF model. The large shape differences between (b) and (c) are mainly due to the different dot sizes and not due to the strain modeling. (d) Hole ground state for $b = 13.6\,\mathrm{nm}$ from an effective mass calculation using the same strain distribution as in (c). The ”quantum numbers” graphically describe the wave-function shapes and originally count the nodal surfaces tangential to (110), $(1\bar{1}0)$, and (001), respectively (see the text on p. 50 and Fig. 4.9).

Electron levels (CM)

For QDs with $b \leq 10\,\mathrm{nm}$ at most one bound electron state is found, assuming the $\mathbf{k}\cdot\mathbf{p}$ parameters from Table 4.2,[3] while for $b = 20\,\mathrm{nm}$ there are as many as 12. The ground state is s-like (see also Fig. 4.10) for all considered dot sizes and its shape is slighty elongated along [110]. The first and second excited, p-like, states $|100\rangle$ and $|010\rangle$ form a non-degenerate pair, the energy separation between them is approximately proportional to b, in agreement with the proportionality between the piezoelectric potential and b. The separation between the average energy of the p-like states and the s-like state decreases as the dot size increases. Intermixing with WL states is not observed, i. e., bound electron wave functions do not extend into the WL.

Hole levels (CM)

Bound hole states are generally confined to the bottom of the pyramid, see Fig. 4.11(b,c), due to the dominant HH contribution in connection with the present strain profile, as discussed on p. 28. Due to the pronounced band-mixing effects discussed in Chapter 3 the shapes of bound holes depend strongly on the QD size, as can be seen by comparing Fig. 4.11(b) ($b = 20.4\,\mathrm{nm}$) and Fig. 4.11(c)[4] ($b = 13.6\,\mathrm{nm}$). The wave functions of excited hole states leak somewhat into the WL. In contrast to single-band effective mass calculations using the same strain distribution, and shown in Fig. 4.11(d), the ground-state wave functions are pronouncedly elongated along $[1\bar{1}0]$ (see Fig. 4.11c), and also the excited states differ significantly from effective mass results [114] with respect to both the wave-function shapes and level separations.

In large QDs the ground-state wave functions are asymptotically torn into two (see Fig. 4.11b), becoming almost degenerate with the first excited level (see Fig. 4.8), thus realizing a geometrical symmetry breaking. In Fig. 4.8 higher excited hole levels for $b > 16\,\mathrm{nm}$ have not been calculated. The dependence of the ground-state energy on the dot size is weaker than for electrons, due to the larger average mass. The energetic splitting between the first and second excited states ($|010\rangle$ and $|020\rangle$) is comparable to that of the electrons ($|100\rangle$ and $|010\rangle$) but not strictly proportional to b anymore, and both states are not p-like either. This underlines that 0D holes behave less classically than electrons because of their greater sensitivity to band-mixing effects caused by the quantum confinement.

---

[3] As shown in Section 3.2.3, four bound CB states can be predicted for such small dots if different material parameters are used (for instance, those from Table 3.1).

[4] The VFF-based wave functions shown in Fig. 4.11(c) to facilitate the comparison with a number of other works look very similar to the CM-based ones.

4.3.2 Impact of input values

Table 4.3 shows some results regarding the influence of variations of selected input parameters, like the InAs/GaAs band offset, on the calculated energy levels. A related study was reported in Ref. [32].

Set	Changed parameters	Unit	Values for GaAs	Values for InAs
1	like no. 2, except			
	voxel size		half lattice constant	
2	**reference set**			
3	like no. 2, except			
	e_{14}	Cm^{-2}	$0.16 \rightarrow 0$	$0.045 \rightarrow 0$
4	like no. 3, except			
	E'_v	meV	$-6920 \rightarrow -6920$	$-6747 \rightarrow -6670$
5	like no. 4, except			
	E_p	eV	$28.0 \rightarrow 22.71$	$22.2 \rightarrow 20.2$
6	like no. 5, except			
	a_c	meV	$-8013 \rightarrow -7170$	
	a_g	meV	$-8233 \rightarrow -8330$	
	b_v	meV	$-1824 \rightarrow -1600$	
	d_v	meV	$-5062 \rightarrow -4230$	$-3600 \rightarrow -3100$
7	like no. 6, except			
	strain (ϵ_{ij})		reduced by 10 %	

Table 4.3: Changes in material parameters made in Fig. 4.12.

Discretization error

The $\mathbf{k}\cdot\mathbf{p}$ material parameters from Table 4.2 are used in sets 1 and 2 of Table 4.3. In set 2 the finite-differences voxel size is the lattice constant, as throughout this part. Set 1 is calculated on a twice as fine grid (8 times more voxels). By comparing the two, the discretization error of the level energies calculated using the grid of set 2 is estimated to be $\leq 10\,\mathrm{meV}$. Since electron and hole energies shift in the same direction, the error in transition energies is even smaller.

Different material parameters

Set 7 is similar to the data used in Ref. [26], and sets 3–6 show the transition from set 2 to set 7 (subsequent sets include the changes of previous ones). In set 3 the piezoelectricity is turned off. In set 4 the band alignment is changed from alignment A (see p. 47) to that calculated in Ref. [58] (called lineup "B") where the unstrained VB offset at the InAs/GaAs heterojunction is 264 meV ($\Delta E_c/\Delta E_g = 0.76$). In sets 5 and 6 the optical matrix parameter and the deformation potentials are changed, respectively, and set 7 takes approximately into account the different strain distributions assumed in Ref. [24, 26].

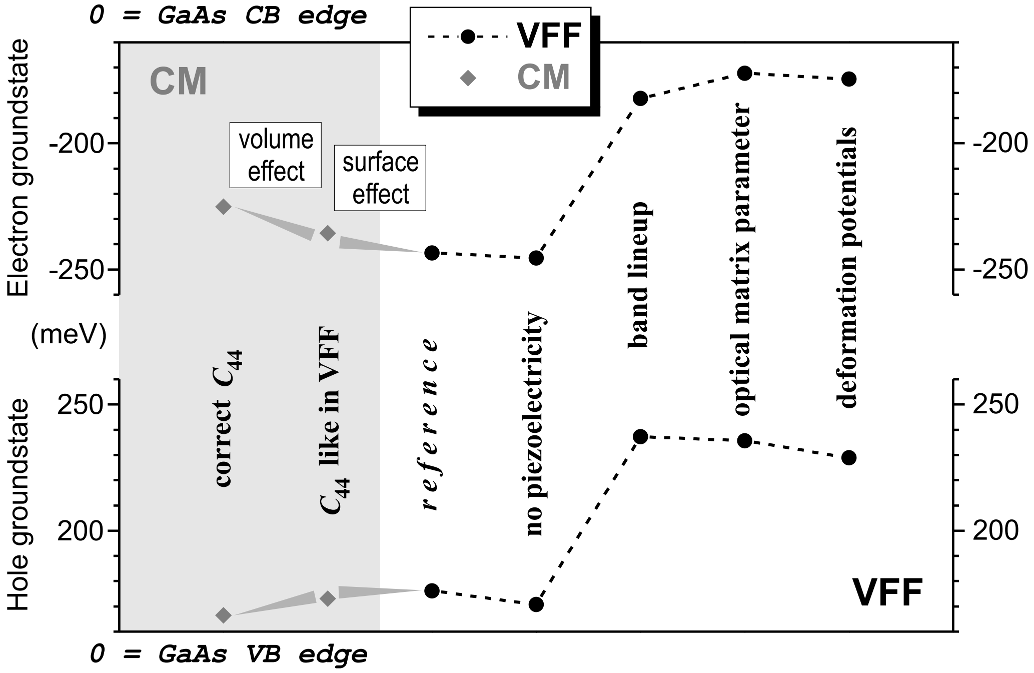


Figure 4.12:

Ground-state energies for $b = 13.6\,\mathrm{nm}$ and different strain models and material parameters, as listed in Tables 4.3 and 4.4. The dotted lines are guides-to-the-eye.

The resulting ground-state energy levels are plotted in Fig. 4.12. The piezoelectricity has little influence on the ground-state *energies*, as also shown in Fig. 3.1(c,d), but this must not distract from its importance for the excited state *wave functions* and the related optical properties like oscillator strengths and anisotropies, as discussed in Chapter 3 and below. The band alignment has little influence on the transition energy. The scaling of strain has little influence on the hole ground-state energy, because $|a_v| \ll |a_c|$ for InAs (see Table 4.2). For smaller optical matrix parameters E_p the transition energy increases which is mainly due to the related effective decrease of the bulk conduction band mass, the holes are much less affected.

	CM	CM[C_{44}^{VFF}]	VFF
E_{000}^{e}	1292.9	1282.4	1274.5
E_{000}^{h}	166.5	173.1	176.2
ΔE	1126.4	1109.2	1098.4

Table 4.4:

Ground-state energies (meV) of electrons and holes and the transition energy in InAs pyramid ($b = 13.6\,\mathrm{nm}$) calculated for three different strain models.

InAs/GaAs band offsets An important parameter for the energy level positions of electrons and holes with respect to the GaAs barrier is the heterojunction band line-up (see also p. 27). Unfortunately, this is one of the least accurately known parameters at the same time. In case that lineup B (see p. 54) were closer to reality than lineup A, the ground-state energies in QDs with $b \geq 13\,\mathrm{nm}$ would shift upwards by $\approx 70\,\mathrm{meV}$ while the increase of kinetic energy of the holes is approximately equal to the decrease of that of the electrons. Thus the ground-state transition energies would increase by less than 15 meV. Some bound electron states would disappear for large dots as compared to the results of Fig. 4.8, and some new bound hole states would appear. The relation between the localization energies of the electron and hole ground states would become almost inverted.

4.3.3 Results for strain according to the VFF model

Table 4.4 shows the ground-state energies of electrons and holes and the direct transition energies calculated using the strain distributions according to the CM, VFF, and CM[C_{44}^{VFF}] models. The differences between CM[C_{44}^{VFF}] and CM are larger than between CM[C_{44}^{VFF}] and VFF (see also Fig. 4.12) indicating that the major part of the difference between CM and VFF is due to the incorrect value of C_{44} in the VFF model rather than its atomistic character. Therefore it is concluded that the CM model gives a more useful description than the (linearized) VFF model. This holds the better the larger the QDs are.

The electron levels are almost rigidly shifted by $-17\,\mathrm{meV}$ as compared to the CM calculation, in agreement with the fact that CM and VFF produce effective electron confinement potentials differing by 20 meV [122]. The hole levels are in average 11 meV stronger localized than according to the CM calculation, otherwise no significant difference is observed. Further details concerning the comparison of the CM and VFF strain models are published in Ref. [31].

4.4 Dipole transitions

Within the dipole approximation the photon absorption cross section [134] of a QD is proportional to

$$M_{ab} = |\,\hat{e} \cdot \mathbf{p}_{ab}\,|^2$$

where $\hat{e}$ is the direction unit vector of the electric field of the linearly polarized, incident light and $\mathbf{p}_{ab}$ is the momentum matrix element between the states $|a\rangle$ and $|b\rangle$. The matrix structure of the $\mathbf{k}\cdot\mathbf{p}$ momentum operator [52]

$$\mathbf{p} = \frac{m_0}{\hbar}\frac{\partial H}{\partial \mathbf{k}} \tag{4.16}$$

(H is the QD Hamiltonian from Eq. (4.14), $\mathbf{k}$ the wave vector, and m_0 the free electron mass) follows from Eq. (2.1):

$$\mathbf{p} = \begin{pmatrix} F & \mathbf{0}_4 \\ \mathbf{0}_4 & -\overline{F} \end{pmatrix} \qquad \text{with} \qquad F = \frac{\partial G}{\partial \mathbf{k}} \quad , \tag{4.17}$$

where $\mathbf{0}_4$ is the 4×4 zero matrix, and the overlining denotes the complex conjugation. The origin of dipole oscillator strength between eight-band $\mathbf{k}\cdot\mathbf{p}$ initial (Ψ_a) and final (Ψ_b) states can be visualized by the following approximation [27, 53] (which is not used in this work, however):

$$\begin{aligned}
\langle\Psi_a|\hat{\mathbf{e}}\cdot\mathbf{p}|\Psi_b\rangle &= \left\langle \sum_{i=1}^{8} u_i\phi_i^{(a)} \middle| \hat{\mathbf{e}}\cdot\mathbf{p} \middle| \sum_{j=1}^{8} u_j\phi_j^{(b)} \right\rangle \\
&= \sum_{i=1}^{8}\sum_{j=1}^{8} \langle u_i\phi_i^{(a)}|\hat{\mathbf{e}}\cdot\mathbf{p}|u_j\phi_j^{(b)}\rangle \\
&= \sum_{i=1}^{8}\sum_{j=1}^{8} \left[\langle u_i|\hat{\mathbf{e}}\cdot\mathbf{p}|u_j\rangle\langle\phi_i^{(a)}|\phi_j^{(b)}\rangle + \langle\phi_i^{(a)}|\hat{\mathbf{e}}\cdot\mathbf{p}|\phi_j^{(b)}\rangle\delta_{ij}\right] \\
&\approx \sum_{i=1}^{8}\sum_{j=1}^{8} \langle u_i|\hat{\mathbf{e}}\cdot\mathbf{p}|u_j\rangle\langle\phi_i^{(a)}|\phi_j^{(b)}\rangle \qquad (4.18) \\
&= \sum_{i=1}^{8}\sum_{j=1}^{8} \langle u_i|e_x p_x + e_y p_y + e_z p_z|u_j\rangle\langle\phi_i^{(a)}|\phi_j^{(b)}\rangle \\
&= \sum_{i=1}^{8}\sum_{j=1}^{8} \begin{pmatrix} e_x\,\langle u_i|p_x|u_j\rangle \\ +e_y\,\langle u_i|p_y|u_j\rangle \\ +e_z\,\langle u_i|p_z|u_j\rangle \end{pmatrix} \langle\phi_i^{(a)}|\phi_j^{(b)}\rangle
\end{aligned}$$

$$
= \quad \begin{array}{l}
e_x \begin{pmatrix} \langle s\uparrow|p_x|x\uparrow\rangle\langle\phi^{(a)}_{s\uparrow}|\phi^{(b)}_{x\uparrow}\rangle \\ +\langle s\downarrow|p_x|x\downarrow\rangle\langle\phi^{(a)}_{s\downarrow}|\phi^{(b)}_{x\downarrow}\rangle \\ +\langle x\uparrow|p_x|s\uparrow\rangle\langle\phi^{(a)}_{x\uparrow}|\phi^{(b)}_{s\uparrow}\rangle \\ +\langle x\downarrow|p_x|s\downarrow\rangle\langle\phi^{(a)}_{x\downarrow}|\phi^{(b)}_{s\downarrow}\rangle \end{pmatrix} + \\
e_y \begin{pmatrix} \langle s\uparrow|p_y|y\uparrow\rangle\langle\phi^{(a)}_{s\uparrow}|\phi^{(b)}_{y\uparrow}\rangle \\ +\langle s\downarrow|p_y|y\downarrow\rangle\langle\phi^{(a)}_{s\downarrow}|\phi^{(b)}_{y\downarrow}\rangle \\ +\langle y\uparrow|p_y|s\uparrow\rangle\langle\phi^{(a)}_{y\uparrow}|\phi^{(b)}_{s\uparrow}\rangle \\ +\langle y\downarrow|p_y|s\downarrow\rangle\langle\phi^{(a)}_{y\downarrow}|\phi^{(b)}_{s\downarrow}\rangle \end{pmatrix} + \\
e_z \begin{pmatrix} \langle s\uparrow|p_z|z\uparrow\rangle\langle\phi^{(a)}_{s\uparrow}|\phi^{(b)}_{z\uparrow}\rangle \\ +\langle s\downarrow|p_z|z\downarrow\rangle\langle\phi^{(a)}_{s\downarrow}|\phi^{(b)}_{z\downarrow}\rangle \\ +\langle z\uparrow|p_z|s\uparrow\rangle\langle\phi^{(a)}_{z\uparrow}|\phi^{(b)}_{s\uparrow}\rangle \\ +\langle z\downarrow|p_z|s\downarrow\rangle\langle\phi^{(a)}_{z\downarrow}|\phi^{(b)}_{s\downarrow}\rangle \end{pmatrix}
\end{array}
$$

$$
= \quad -\mathrm{i}\,\frac{m_0}{\hbar}\,P \left[\begin{array}{l}
e_x \begin{pmatrix} \langle\phi^{(a)}_{s\uparrow}|\phi^{(b)}_{x\uparrow}\rangle - \langle\phi^{(a)}_{x\uparrow}|\phi^{(b)}_{s\uparrow}\rangle \\ +\langle\phi^{(a)}_{s\downarrow}|\phi^{(b)}_{x\downarrow}\rangle - \langle\phi^{(a)}_{x\downarrow}|\phi^{(b)}_{s\downarrow}\rangle \end{pmatrix} + \\
e_y \begin{pmatrix} \langle\phi^{(a)}_{s\uparrow}|\phi^{(b)}_{y\uparrow}\rangle - \langle\phi^{(a)}_{y\uparrow}|\phi^{(b)}_{s\uparrow}\rangle \\ +\langle\phi^{(a)}_{s\downarrow}|\phi^{(b)}_{y\downarrow}\rangle - \langle\phi^{(a)}_{y\downarrow}|\phi^{(b)}_{s\downarrow}\rangle \end{pmatrix} + \\
e_z \begin{pmatrix} \langle\phi^{(a)}_{s\uparrow}|\phi^{(b)}_{z\uparrow}\rangle - \langle\phi^{(a)}_{z\uparrow}|\phi^{(b)}_{s\uparrow}\rangle \\ +\langle\phi^{(a)}_{s\downarrow}|\phi^{(b)}_{z\downarrow}\rangle - \langle\phi^{(a)}_{z\downarrow}|\phi^{(b)}_{s\downarrow}\rangle \end{pmatrix}
\end{array} \right] . \qquad (4.19)
$$

Here, u_i denote the eight Bloch functions and ϕ_i the corresponding envelope functions. The ”$\approx$” sign in Eq. (4.18) refers to the basic envelope function approximation according to which each envelope ϕ_i is much slower varying than its corresponding Bloch function u_i , hence $\nabla\phi_i$ terms can be neglected aside of ∇u_i terms. As can be seen from Eq. (4.19) oscillator strength arises, if applicable, from significant overlap between the s-type envelopes of the initial state and the p-type envelopes of the final state. Interband transitions usually yield larger values of M_{ab} than intraband transitions, given the same orbital overlap between the probability densities of the initial and final states.

By the use of Eq. (4.19) structural information about the QD would enter the momentum matrix elements only via the wave functions of the initial and final states, while the momentum operator **p** is reduced to a scalar optical matrix element P. Such a curtailment is in obvious contrast to Eq. (4.17) where the matrix F contains more information about the heterostructure than P: Eq. (4.19) does not support a consistent definition of **p** with respect to H [52]. This shortcoming is avoided in the present work by using Eq. (4.16) instead:

$$
\begin{aligned}
\mathbf{p}_{ab} &= \frac{m_0}{\hbar}\,\langle a|\frac{\partial H}{\partial \mathbf{k}}|b\rangle \quad , \qquad (4.20) \\
M_{ab} &= \frac{{m_0}^2}{\hbar^2}\left|\hat{e}\cdot\langle a|\frac{\partial H}{\partial \mathbf{k}}|b\rangle\right|^2 \quad ,
\end{aligned}
$$

$|a\rangle$ and $|b\rangle$ represent the eight-band $\mathbf{k}{\cdot}\mathbf{p}$ envelope function vectors $(\phi_1, \ldots, \phi_8)$ as in Eq. (4.15). Eq. (4.20) provides an appropriate treatment of the structural, $\mathbf{k}$-, and strain dependence of $\mathbf{p}$ and yields values of M_{ab} consistent with Eq. (4.14).

In case that $P(\mathbf{r})$ has a constant value for $\mathbf{r}$ inside the QD Eq. (4.19) suggests a normalization of $\mathbf{p}_{ab}$ in order to eliminate the bulk contribution (represented by the Bloch function part, $-\mathrm{i}m_0 P/\hbar$) and to isolate the contribution from the confined wave functions, i. e. their orbital overlap.[5] This is achieved by scaling M_{ab} to the optical matrix parameter E_p (see p. 9) of the QD material (InAs), yielding a dimensionless, scaled matrix element (SME)

$$I \;=\; \frac{2}{m_0\,E_p}\,M_{ab} \;=\; \frac{1}{P^2}\left|\hat{e}\cdot\langle a|\frac{\partial H}{\partial \mathbf{k}}|b\rangle\right|^2 \tag{4.21}$$

which is analogous to the overlap integral $|\langle a|b\rangle|^2$ determining the optical selection rules within the single-band effective-mass approximation [55]. However, in the eight-band $\mathbf{k}{\cdot}\mathbf{p}$ model I depends on $\hat{e}$ and describes, additionally, the polarization anisotropy while in the single-band approximation the anisotropy must be examined independently. To model the interaction of a QD with diffuse or unpolarized light, I has to be averaged (absorption) or integrated (spontaneous emission), respectively, over the respective directions of $\hat{e}$.

The SME I is not identical with the oscillator strength which is obtained from Eq. (4.21) by replacing E_p with the transition energy $E_{ab} = |E_a - E_b| = \hbar\omega_{ab}$ [42,88] and thus proportional to I by the factor E_p/E_{ab}. The oscillator strength, being proportional to absorption coefficients [28,42,53,88] or cross sections [134], introduces to those a universal, explicit dependence on $1/E_{ab}$ which is responsible for a significantly enhanced interaction of the QD with mid-infrared photons as compared to visible or near-infrared photons (given the same irradiation intensities and M_{ab} values). This effect is due to the ω_{ab}-dependent interrelations between the optical intensity S, the photon flux Φ, and the magnitude A of perturbation to the quantum-mechanical two-level system, as can be seen considering the light field classically in the Coulomb gauge [135]: The electromagnetic wave is completely represented by the vector potential $\mathbf{A}$, and $A \propto |\mathbf{A}|^2$ is proportional to Φ/ω while $S \propto \omega\Phi$. The photon loss is $\Delta\Phi \propto M_{ab}A$ so that the absorption $\alpha \propto \Delta\Phi/\Phi$ is proportional to M_{ab}/ω_{ab}. This elementary ω-dependence of α is neglected throughout this work using I to characterize the impact of quantum confinement effects on the optical properties of QDs, so that the SME is considered rather than the oscillator strength.

---

[5] The penetration of the confined particles into the surrounding barrier material, usually having a different value of P, is neglected here.

4.4.1 Time reversal symmetry

In absence of magnetic fields all single-particle levels in a QD are two-fold[6] degenerate by time reversal symmetry [87, 88]: If Ψ_1 according to Eq. (4.15) is an eigenvector to Eq. (4.14) then

$$\Psi_2 = \begin{pmatrix} \mathbf{0}_4 & -\mathbf{1}_4 \\ \mathbf{1}_4 & \mathbf{0}_4 \end{pmatrix} \overline{\Psi_1} \tag{4.22}$$

is another eigenvector belonging to the same energy and being orthogonal to Ψ_1 ($\mathbf{1}_4$ is the 4×4 identity matrix). Due to Eq. (4.22) the orthogonality relation $\langle \Psi_1 | \Psi_2 \rangle = 0$ holds pointwise for each spatial coordinate $\mathbf{r}$: The wave-function product

$$\Psi_1(\mathbf{r}) \cdot \Psi_2(\mathbf{r}) \equiv 0 \tag{4.23}$$

is a scalar product between two orthogonal, eight-dimensional envelope vectors of the type Eq. (4.15).

In the presence of magnetic fields the time reversal symmetry is destroyed and the associated single-particle level degeneracy lifted, thus Ψ_1 and Ψ_2 belong to different energies then.

The 2×2 block matrix in Eq. (4.22) "flips" the spin between the two basis vectors Ψ_1 and Ψ_2 of the degenerate eigenspace. However, due to the presence of Γ in Eq. (2.1) the spin-up ($\uparrow$) and spin-down ($\downarrow$) projections of the wave functions $\Psi_{1,2}$ are inevitably coupled, as in the bulk case [48], hence the presence of SO interaction in the bulk materials ($\Delta_0 \neq 0$) prevents spin conservation also in the QD: The spin yields no good quantum number for the single-particle states.

Moreover – unlike the bulk case – the band mixing effects discussed in Section 3.2.1 prevent a strict classification of 0D CB and VB states according to their angular momenta, i. e., neither the bound electrons are $|\pm 1/2\rangle$ states, nor the (topmost) bound holes are $|\pm 3/2\rangle$ states.

At time reversal symmetry the momentum matrix elements for single-particle transitions (primarily electron-electron or hole-hole) are obtained by incoherent averaging over the degenerate eigenspaces a and b, giving

$$\begin{aligned} |\,\hat{e} \cdot \mathbf{p}_{ab}\,|^2 &= \frac{1}{4}\left(|\hat{e} \cdot \mathbf{p}_{a_1 b_1}|^2 + |\hat{e} \cdot \mathbf{p}_{a_2 b_1}|^2 + |\hat{e} \cdot \mathbf{p}_{a_1 b_2}|^2 + |\hat{e} \cdot \mathbf{p}_{a_2 b_2}|^2 \right) \\ &= \frac{1}{2}\left(|\hat{e} \cdot \mathbf{p}_{a_1 b_1}|^2 + |\hat{e} \cdot \mathbf{p}_{a_1 b_2}|^2 \right) \qquad \text{since} \\ \mathbf{p}_{a_2 b_1} &= \overline{\mathbf{p}_{a_1 b_2}} \quad \text{and} \\ \mathbf{p}_{a_2 b_2} &= -\overline{\mathbf{p}_{a_1 b_1}} \qquad , \end{aligned} \tag{4.24}$$

---

[6] Accidental geometrical degeneracies may additionally enhance the multiplicity of an eigenvalue.

where $|a_1\rangle$ and $|a_2\rangle$ ($|b_1\rangle$ and $|b_2\rangle$) satisfy Eq. (4.22). The resulting dependence of I on the direction of $\hat{e}$ is shown in Fig. 5.2(a) for the QD size $b = 17\,\mathrm{nm}$.

Eq. (4.24) cannot be reduced further since $\Delta_0 \neq 0$ prevents a decoupling of the orbital and spin parts of $|b_1\rangle$ and $|b_2\rangle$. However, in absence of strain[7] and band mixing $\mathbf{p}_{a_1 b_1}$ and $\mathbf{p}_{a_2 b_2}$ would, by convention, correspond to the dipole-allowed (total angular momentum difference $\Delta M = \pm 1$) transitions $|-1/2\rangle \rightarrow |+3/2\rangle$ and $|+1/2\rangle \rightarrow |-3/2\rangle$, respectively, while $\mathbf{p}_{a_1 b_2}$ and $\mathbf{p}_{a_2 b_1}$ would vanish since being related to the dipole-forbidden ($\Delta M = \pm 2$) transitions $|-1/2\rangle \rightarrow |-3/2\rangle$ and $|+1/2\rangle \rightarrow |+3/2\rangle$, respectively. Hence, M_{ab} calculated from Eq. (4.24) yields a SME I corresponding to the half of the single-band overlap integral, $|\langle a|b\rangle|^2/2$.

---

[7] As discussed in Section 3.2.1 and Ref. [39] the VB edge in the uppermost region of pyramidal QDs is defined by the LH band, while near the bottom by the HH band, because of the particular strain distribution. In other cases tensile strain may push the LH band edge atop of the HH band edge. In such situations selection rules based on the total angular momentum are not useful.

Chapter 5

Few-Particle States in Quantum Dots

A demanding challenge in the modeling of 0D systems arises from the fact that the energy level structure of a QD depends on the population of its levels with charge carriers, i. e., renormalization effects play an elementary role in QD heterostructures. The considerations of Chapter 4 apply strictly to empty QDs only, holding for the first charge carrier occupying the dot. As soon as more than one electron or hole is confined in the QD, the impact of (i) direct[1] Coulomb interaction, (ii) exchange effects [87], and (iii) correlation lead to the formation of distinct many-body states with different energies and transition characteristics.

The balance between the three effects varies dependent on the type of many-body state, i. e. the number of confined particles and their charges, and on the confinement profile, in particular the size and symmetry of the QD. In the strong confinement regime many-body effects act as a perturbation to the confinement and excitons are weakly correlated (see p. 35). The direct Coulomb interaction yields the largest contribution to the total energy of an exciton or biexciton state as compared to the sum of energies of particles in "corresponding" single-particle states. The parentheses indicate that a strict correspondence between a many-body state and certain single-particle states, being understood as initial states for the formation of the correlated many-body state, does not exist. This inexistence is caused by the correlation of particles in a many-body state which can be visualized in different ways:

- Certain "initial" single-particle states *deform* themselves under the influence of Coulomb interaction in order to minimize the total energy of the resulting many-body state. Thus correlation means deformation of single-particle wave functions. It depends only on the extent of this deformation whether the "initial" states can be identified from a given many-body state. The weak correlation present in the strong confinement regime often allows such an identification, see also p. 72.

---

[1] "Direct" denotes the classical electrostatic forces between *classical* electrons and holes, i. e., not Fermions.

- If the empty QD possesses a sufficiently large number of bound CB and VB states to serve as a, for practical purposes, complete orthonormal basis the deformation of each "initial" single-particle state can be expanded into a series of excited states of that particle. The many-body state is thus expanded into *configurations* built from the ground- and excited single-particle states, and the correlation manifests itself by the significance of more than one configuration within this expansion. In this picture, correlation means configuration mixing.

The method related to this second notion of correlation is the configuration interaction (CI) method. It is elegant, though not necessary [89,136–138], to use the eigenstates of the empty QD as basis for many-body expansions in the strong confinement regime [139–144], since the obtained basis configurations are already similar to the weakly correlated many-body states, thus providing a quickly converging approximation. The suitability of this choice implicitly exploiting the perturbational character of many-body effects in the strong confinement regime is limited in several directions, one of which is the assumption of strong confinement itself:

- In the case of small QDs the number of existing eigenstates may be insufficient for a convergent expansion. In the limiting case where only one bound CB and VB state exist, correlation cannot be modeled by a CI approach neglecting continuum states.

- In the case of a type-II band alignment at the QD boundary surface, like in the Ga(As)Sb/GaAs material system [1,145], only holes (e. g.) are bound in the QD, while electrons are only localized by dots charged with holes. Electrically neutral many-body states, like excitons [136,146,147], in such dots cannot be modeled by a CI method relying on a single-particle eigenstate basis.[2]

- In the intermediate (or even weak) confinement regime the correlation may be so pronounced that an exceedingly large configuration basis becomes necessary [138].

---

[2] However, a variational CI method using non-orthogonal basis functions depending on variational parameters is applicable to calculate type-II excitons in the Hartree approximation [136].

These limitations can be overcome by self-consistent calculations in the frame of the local density approximation (LDA) [31, 148–153] or the local spin density approximation (LSDA) [20,22,154–159]. Both the self-consistent calculations and the CI method offer the advantage over variational methods [106, 147, 160–164] that they (i) do not minimize the total energy of the ground state over a restricted, intuitive, parametric subspace, and (ii) readily yield excited many-body states.

5.1 LDA in the Hartree approximation

The self-consistent LDA[3] is demonstrated calculating the exciton, trion, and biexciton ground states in the Hartree approximation [87], i. e., neglecting the exchange effects.

The **exciton ground state** is calculated using the ansatz of a separable exciton wave function $\Psi_X(\mathbf{r}_e\,,\mathbf{r}_h) = \Psi_e(\mathbf{r}_e)\Psi_h(\mathbf{r}_h)$ whose constituing single-electron (Ψ_e) and -hole (Ψ_h) states are determined self-consistently [31]:

$$[H + V_e]\Psi_h = \tilde{E}_h\Psi_h \quad , \quad [H + V_h]\Psi_e = \tilde{E}_e\Psi_e \quad , \tag{5.1}$$

$$-e|\Psi_e|^2 = \epsilon_0\nabla(\epsilon_s\nabla V_e) \quad , \quad e|\Psi_h|^2 = \epsilon_0\nabla(\epsilon_s\nabla V_h) \quad , \tag{5.2}$$

H is the empty-dot Hamiltonian from Eq. (4.14). Thus the deformation of single-particle wave functions is explicitly calculated, relating the procedure to the first of the above notions of correlation. This approach is limited to the many-body ground state but, on the other hand, requires only the single-particle ground states, which renders it appropriate to the treatment also of the smallest dots in Fig. 4.8. Like Eq. (4.12), Eq. (5.2) includes the image charge effect. Due to the Hartree approximation used here the exciton ground state is four-fold degenerate. This degeneracy can be split (i) by applying magnetic fields destroying the time reversal symmetry and lifting the degeneracy of the single-particle levels, and (ii) by exchange effects leaving the single-particle level structure unaltered. The treatment of exchange effects requires to take into account the Fermion character of electrons and holes, as accomplished by the Hartree-Fock approximation.

Using the notations

$$C = (\langle\Psi_h|V_e|\Psi_h\rangle - \langle\Psi_e|V_h|\Psi_e\rangle)/2 \tag{5.3}$$

$$C_e = \langle\Psi_e|V_e|\Psi_e\rangle \tag{5.4}$$

$$C_h = -\langle\Psi_h|V_h|\Psi_h\rangle \tag{5.5}$$

---

[3] A more correct, though less common, terminology refers to these calculations as self-consistent mean field theory [73]. The idea to call the present treatment LDA is based on the mesoscopic notion of quasi-particles being represented by the envelope functions, neglecting the true atomistic properties of the wave functions.

b (nm)	$E^{\mathrm{e}}_{\mathrm{coul}}$	${E_X}^{\mathrm{LDA}}$	${E_X}^{\mathrm{CI}}$	p (direct)	p (exciton)
10.2	24.8	26.1	27.7	0.056	0.052
13.6	24.3	21.7	23.7	0.098	0.089
17.0	20.7	15.6	17.5	0.18	0.15
20.4	17.8	11.0	12.5	0.26	0.16

Table 5.1:

Size (b) dependence of the electron ground-state Coulomb charging energy ($E^{\mathrm{e}}_{\mathrm{coul}}$), the exciton ground-state binding energy (E_X), and of the polarization anisotropy index p for hypothetical direct recombination (direct) and excitonic recombination (exciton), self-consistently calculated in the Hartree LDA, and E_X calculated using CI in the Hartree approximation. All energies are given in meV, see the text for their discussion. p is defined by Eq. (6.2) on p. 86.

the exciton binding energy E_X and the exciton recombination energy $E^{(X)}$ in the Hartree approximation are

$$E^{(X)} = C + \tilde{E}_e - \tilde{E}_h \quad , \tag{5.6}$$

$$E_X = E_e - E_h - E^{(X)} \quad , \tag{5.7}$$

where E_e and E_h denote the single-electron and -hole energy, respectively (being eigenvalues of H). The binding energy E_X as a function of the QD size b is listed in Table 5.1. As a consequence of the separable ansatz, the two-particle character of the exciton is not appropriately captured by this approach, i. e., the treatment of correlation is still restricted.[4] However, the effect of this constraint on the total energy is < 0.5 meV as determined from comparison with CI calculations in the Hartree approximation also shown in Table 5.1. The actual differences between both calculations range from 1.5 meV to 2.0 meV whereof 1.5 meV are caused by different numerical procedures used for solving Eq. (5.2) in the LDA and CI methods, respectively, i. e. the significant part of the differences is < 0.5 meV.

The radiative recombination of the exciton can be modeled inserting the self-consistent states $|\Psi_e\rangle$ and $|\Psi_h\rangle$ from Eq. (5.1) as $|a\rangle$ and $|b\rangle$ into Eq. (4.20). Though at self-consistency the eigenspaces a and b are not orthogonal any more (since they belong to separate Hamiltonians differing in their confinement potential parts) $\mathbf{p}_{ab}$ is not affected because $\partial H/\partial \mathbf{k}$ does not depend on scalar (electrostatic) potentials. The oscillator strength obtained this way is a lower bound for the true excitonic oscillator strength because the two-particle character of the exciton was ignored by the separable wavefunction ansatz. Accordingly, an upper bound is obtained for the optical anisotropy p defined by Eq. (6.2). Nevertheless, the reduction of the excitonic anisotropy as compared to the direct recombination becomes clear also from the Hartree LDA calculation, see Table 5.1.

---

[4] According to a more strict, though less common, notion self-consistent calculations assuming separable many-body wave functions do not describe correlation at all, since correlation then is defined as the *difference* between self-consistent mean field results and a true many-body wave function as obtained by, e. g., full CI or quantum Monte-Carlo calculations [73]. This definition of correlation agrees better with that known from statistics, but is less graphical.

The **two-electron ground state** is obtained in an analogous manner,

$$[H + V_e]\Psi_e = \tilde{E}_e\Psi_e \quad , \tag{5.8}$$

with V_e according to Eq. (5.2). The electron ground-state Coulomb charging energy is $E^{\mathrm{e}}_{\mathrm{coul}} = \tilde{E}_e - E_e$ and listed in Table 5.1.

The ground state of the **negative trion** (an exciton charged with one excess electron) is obtained from

$$[H + 2V_e]\Psi_h = \tilde{E}_h\Psi_h \qquad , \qquad [H + V_e + V_h]\Psi_e = \tilde{E}_e\Psi_e$$

together with Eq. (5.2) and yields a recombination energy of

$$E^{(T^-)} = C + \tilde{E}_e - \tilde{E}_h \qquad .$$

The ground state of the **positive trion** (an exciton charged with one excess hole) is obtained from

$$[H + V_e + V_h]\Psi_h = \tilde{E}_h\Psi_h \qquad , \qquad [H + 2V_h]\Psi_e = \tilde{E}_e\Psi_e$$

together with Eq. (5.2) and yields a recombination energy of

$$E^{(T^+)} = C + \tilde{E}_e - \tilde{E}_h \qquad .$$

The **biexciton ground state** is obtained from

$$[H + 2V_e + V_h]\Psi_h = \tilde{E}_h\Psi_h \qquad , \qquad [H + V_e + 2V_h]\Psi_e = \tilde{E}_e\Psi_e$$

together with Eq. (5.2) and has the energy

$$E^{(XX)} = 2\left(\tilde{E}_e - \tilde{E}_h\right) + 4C - C_e - C_h \qquad .$$

For pyramidal InAs/GaAs QDs with a base length of 17 nm the following energies are obtained:

$$\begin{array}{lcll}
E^{(X)} & = & 1027.0\ \mathrm{meV} & \\
E^{(T^-)} & = & 1031.2\ \mathrm{meV} & \\
E^{(T^+)} & = & 1027.5\ \mathrm{meV} & \\
E^{(XX)} & = & 2058.6\ \mathrm{meV} & \\
E^{(XX)} - E^{(X)} & = & 1031.6\ \mathrm{meV} & \text{(biexciton recombination energy)} \\
2E^{(X)} - E^{(XX)} & = & -4.5\ \mathrm{meV} & \text{(biexciton binding energy).}
\end{array}$$

Except from the smallest QDs with $b = 10\,\mathrm{nm}$ the InAs/GaAs dots considered in Chapter 4 possess enough single-particle energy levels to enable a CI treatment of the many-body effects. A comparison with CI calculations in the Hartree-Fock approximation shows that the Hartree LDA yields reasonable approximations for the exciton ground state and the negative trion ($\leq 1\,\mathrm{meV}$ significant[5] error for the 17 nm pyramid). In contrast, the ground-state energies of the positive trion and the biexciton both are overestimated by $\sim 6\,\mathrm{meV}$, suggesting that the *exchange*-correlation effect (being neglected in the Hartree LDA) of the holes is more pronounced than that of the electrons. The biexciton binding energy is underestimated by $\sim 3\,\mathrm{meV}$ only, since the errors in the calculations of $E^{(X)}$ and $E^{(XX)}$ partly compensate each other.

5.2 Configuration interaction calculations

In cases where the CI method is applicable to QD heterostructures it has the advantage over the LDA (LSDA) method that it requires no (spin-polarization dependent) parametrization for the exchange-correlation potential [22] but can treat both the spin and the Fermion character of particles exactly. Especially in connection with EP theory [144] or the eight-band $\mathbf{k}\cdot\mathbf{p}$ model (this chapter) as underlying bandstructure calculation methods the long-range component of the electron-hole exchange effect, which dominates over the short-range component in the case of Wannier excitons [87,88,165], can be calculated directly from the single-particle wavefunctions without further parameters. This advantage is not available if the bulk CBs and VBs are decoupled in the underlying bandstructure calculation, as in the case of single-band effective-mass based solutions of the Hartree-Fock equation in QDs [106,108,138,141,161,163,166–170]. Therefore, the CI method is prefered in this work for calculating the exciton and biexciton level structure in not too small QDs, using pyramidal InAs/GaAs QDs with 17 nm base length as an example.

The principal characteristic of the CI method is that it takes into account many-body correlation by employing a parameter-free, non-separable ansatz for a true many-body wave function. The ansatz is constructed as linear combination of all configurations that can be built from a given set of single-particle eigenstates. These configurations form a combinatorial, orthogonal basis into which the many-body Hamiltonian is expanded. The many-body (eigen)states are obtained by diagonalizing that Hamiltonian. In the Hartree approximation [87], neglecting the Fermion character of the single particles and thus exchange effects, the configuration basis is obtained as cartesian product of the single-particle eigenstates. In the Hartree-Fock approximation [87], including the Pauli principle (and thus exchange effects), the basis is obtained by converting each element of the cartesian product into a Slater determinant, thus using only antisymmetric basis states.

---

[5] See p. 64.

5.2.1 Fermion character and spin

If many-body states consisting of n_e electrons c_i and n_h holes v_j shall be calculated on the basis of the lowest N_e bound CB states and the topmost N_h bound VB states of Eq. (4.14), the CI basis contains

$$N = \binom{N_e}{n_e} \cdot \binom{N_h}{n_h} \tag{5.9}$$

elements. These are either distinct configurations

$$|\, c_1(\mathbf{r}_1) \, \ldots \, c_{n_e}(\mathbf{r}_{n_e}) \, v_1(\mathbf{r}_{n_e+1}) \, \ldots \, v_{n_h}(\mathbf{r}_{n_e+n_h}) \,\rangle$$

(when using the Hartree approximation) or Slater determinants thereof,

$$\frac{1}{\sqrt{(n_e+n_h)!}} \begin{vmatrix} c_1(\mathbf{r}_1) & \ldots & c_{n_e}(\mathbf{r}_1) & v_1(\mathbf{r}_1) & \ldots & v_{n_h}(\mathbf{r}_1) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ c_1(\mathbf{r}_{n_e+n_h}) & \ldots & c_{n_e}(\mathbf{r}_{n_e+n_h}) & v_1(\mathbf{r}_{n_e+n_h}) & \ldots & v_{n_h}(\mathbf{r}_{n_e+n_h}) \end{vmatrix} ,$$

when using the Hartree-Fock approximation. The $\mathbf{r}_i$ are spatial coordinates while the spin parts of the states are accommodated in the wave functions c_i and v_i: A separation of the orbital and spin parts of the 0D wave functions is – like in the bulk case – prevented by the SO interaction in the bulk materials ($\Delta_0 \neq 0$). Furthermore – unlike the bulk case – the pronounced band mixing in strongly confining QDs merges the bulk HH and LH bands, thus suspending even total angular momentum selection rules (see Secion 4.4.1). It follows that any treatment of many-body fine-structure effects in QDs based on strict selection rules for the total angular momentum [106,108,109,138,142,161,163,166–168,170] yields inaccurate predictions of level degeneracies, ignoring the band mixing characteristic to 0D structures. However, the basic pattern of the energetic exciton structure is governed by the total angular momentum while the band mixing introduces a perturbation, as shown below.

In the presence of magnetic fields the c_i and v_i belong all to different energies[6] so that exciton level degeneracies are lifted even in absence of band mixing. The CI approach is not changed by this in any respect [171].

The above antimetrization of the CI basis vectors in case of the Hartree-Fock approximation requires the orthogonality of all single-particle states $c_1, \ldots, c_{n_e}$,

---

[6] Apart from possible accidental degeneracies.

$v_1, \ldots, v_{n_h}$, otherwise the Slater determinants do not correctly represent (antisymmetric) Fermion states. If both the CB states c_i and the VB states v_j are obtained from the same Hamiltonian H, as in the case of EP or eight-band $\mathbf{k}\cdot\mathbf{p}$ theory, they are orthogonal by default and generate an orthonormal basis for both Hartree and Hartree-Fock calculations. In contrast, bandstructure models decoupling the bulk CBs and VBs require additional parameters ("exchange strength constant") to treat the "exchange interaction" between CB and VB states [138, 161, 163, 167]. In the EP and eight-band $\mathbf{k}\cdot\mathbf{p}$ models, the 0D CB and VB states are fully consistent and require solely an external assignment of their charge. This renders the eight-band $\mathbf{k}\cdot\mathbf{p}$ model superior to the six-band $\mathbf{k}\cdot\mathbf{p}$ [172,173], four-band $\mathbf{k}\cdot\mathbf{p}$ [24,174], and single-band models for the treatment of excitonic and biexcitonic properties in QDs.

5.2.2 Dielectric screening

The basis vectors for exciton calculations ($n_e = 1$, $n_h = 1$) are of the type $|\, c_1 v_2 \,\rangle$ when using the Hartree approximation and $|\, c_1 v_2 - c_2 v_1 \,\rangle$ when using the Hartree-Fock approximation. They differ by the term $|\, c_2 v_1 \,\rangle$ which sometimes is called "exchange term" while the Hartree basis vector $|\, c_1 v_2 \,\rangle$ is refered to as "direct Coulomb term". It would be tempting to screen the two related matrix elements by different dielectric constants [175] when calculating the direct Coulomb and exchange energies, in order to take into account the different *asymptotic* screening characteristics [4, 144, 176] of the long-ranged direct Coulomb interaction and the short- or long-ranged exchange effects. However, such an approach ignores the common origin of both matrix elements which describe one and the same interaction (Coulomb) between Fermion particles. Within the current frame of an eight-band $\mathbf{k}\cdot\mathbf{p}$ theory based CI treatment it would be inappropriate for further two reasons:

- The envelope function approach is *a priori* questionable when applied to Frenkel excitons [87], for which the short-range exchange is important. *Vice versa*, the modeling of Wannier excitons, for which the $\mathbf{k}\cdot\mathbf{p}$ approach is better suited, requires the same long-range screening for both the direct Coulomb and exchange effects [144, 177]. In addition, only the long-range exchange is sensitive to the shape of the QD [161].
- The associated, implicit loss of antisymmetry of the basis states would cause severe complications of the CI method.

Hence, the CI calculations presented in this work employ only long-range screening, applying the bulk dielectric constants of the constituing materials to all Coulomb matrix elements (i. e., the "direct Coulomb" and the "exchange terms") and taking into account image charge effects. A possibility to improve the treatment of dielectric screening in a CI compatible fashion is to use a microscopic dielectric function depending on both the individual electron and hole positions, as well as on their distance [81, 144], thus unifying the "long range" and "short range" properties. However, this is proprietary to microscopic approaches, e. g. pseudopotential calculations.

5.2.3 Computational expense

The CI Hamiltonian rank for the calculation of excitons (X) is $N_X = N_e N_h$ and that for biexcitons (XX) ($n_e = 2$, $n_h = 2$) is $N_{XX} = N_e N_h (N_e - 1)(N_h - 1)/4$ according to Eq. (5.9), regardless of whether exchange effects are included. In the case of the exciton the computational expense for evaluating one Hamiltonian matrix element is doubled when advancing from the Hartree approximation to the Hartree-Fock approach, due to the additional inclusion of the exchange term. In the biexciton case the replacement of the Hartree basis vectors

$$| \, c_1(\mathbf{r}_1) \, c_2(\mathbf{r}_2) \, v_1(\mathbf{r}_3) \, v_2(\mathbf{r}_4) \; \rangle$$

by antisymmetric Hartree-Fock basis vectors

$$\frac{1}{\sqrt{24}} \begin{vmatrix} c_1(\mathbf{r}_1) & c_2(\mathbf{r}_1) & v_1(\mathbf{r}_1) & v_2(\mathbf{r}_1) \\ c_1(\mathbf{r}_2) & c_2(\mathbf{r}_2) & v_1(\mathbf{r}_2) & v_2(\mathbf{r}_2) \\ c_1(\mathbf{r}_3) & c_2(\mathbf{r}_3) & v_1(\mathbf{r}_3) & v_2(\mathbf{r}_3) \\ c_1(\mathbf{r}_4) & c_2(\mathbf{r}_4) & v_1(\mathbf{r}_4) & v_2(\mathbf{r}_4) \end{vmatrix}$$

enhances the computational load for each matrix element by a factor 24. In addition, the expense for setting up the whole Hamiltonian, i. e. calculating all – possibly different[7] – matrix elements, scales with $\sim N^2/2$ (the expense for diagonalizing the matrix is negligible). Since the inclusion of the Pauli principle is mandatory for the modeling of biexcitons, the CI calculations of exciton and biexciton (in the Hartree-Fock approximation) differ with respect to the number of algebraic terms by a factor

$$\frac{3}{4} \left[\, (N_e - 1) \, (N_h - 1) \, \right]^2 \qquad ,$$

i. e. a factor ~ 1500 if 3 CB levels ($N_e = 6$) and 5 VB levels ($N_h = 10$) are included. The biexciton (exciton) CI Hamiltonian then has a rank of $N_{XX} = 675$ ($N_X = 60$) and contains 63450 (1830) non-zero matrix elements, whose calculation can be reduced by *complete economization* to the evaluation of 136 (136) wave-function products, 166050 (3660) volume integrations, and 135 (81) solutions of the 3D complex Poisson equation for the QD heterostructure including image charge effects. Thus the total economization of the matrix-element calculation reduces the expense factor between the exciton and biexciton treatments from ~ 1500 to ~ 50.

---

[7] In presence of magnetic fields no symmetries are available.

Nevertheless, the example illustrates the combinatorial progression of the calculation expense related to the exact[8] treatment of many-body effects and explains why the CI method, even when employing simplified bandstructure models in connection with simplified real-structure models of the dot, is presently limited to the treatment of *two* biexcitons [138]. This has to be compared to the present work treating *one* biexciton *including* strain, band mixing, and SO interaction.

Systems of more than two biexcitons presently require different approaches, like stochastic diagonalization [178] or quantum Monte-Carlo methods [179].

5.2.4 Exciton in the Hartree approximation

The Coulomb interaction between the electron and hole forming an exciton is calculated expanding the exciton wave function $\Psi^{(X)}$ into a cartesian product basis built from all available 0D electron ($c_a = \Psi_a^{(c)}$) and hole states ($v_b = \Psi_b^{(v)}$):

$$\Psi^{(X)} = \sum_{a,b} \xi_{ab}\, \Psi_a^{(c)} \Psi_b^{(v)} \quad , \tag{5.10}$$

$$\mathbf{p}^{(X)} = \sum_{a,b} \xi_{ab}\, \mathbf{p}_{ab} \quad . \tag{5.11}$$

$\mathbf{p}^{(X)}$ corresponds to the dipole term of the interaction of $\Psi^{(X)}$ with light and the $\mathbf{p}_{ab}$ are given by Eq. (4.20). Since exchange effects are neglected and no magnetic field is present each excitonic energy level $|n\rangle$ possesses a four-dimensional orthonormal basis due to time reversal symmetry

$$\{\Psi_{n,1}, \Psi_{n,2}, \Psi_{n,3}, \Psi_{n,4}\}^{(X)} \quad .$$

The momentum matrix element of such a degenerate energy level with light having an electric field vector parallel to the unit vector $\hat{e}$ is obtained by incoherent averaging over the eigenspace

$$|\, \hat{e} \cdot \mathbf{p}_n^{(X)} \,|^2 = \frac{1}{4} \left(|\hat{e} \cdot \mathbf{p}_{n,1}^{(X)}|^2 + |\hat{e} \cdot \mathbf{p}_{n,2}^{(X)}|^2 + |\hat{e} \cdot \mathbf{p}_{n,3}^{(X)}|^2 + |\hat{e} \cdot \mathbf{p}_{n,4}^{(X)}|^2 \right) \quad , \tag{5.12}$$

analogous to Eq. (4.24). The resulting dependence of the oscillator strength on the E-field direction is shown in Fig. 5.2(b) for the QD size $b = 17\,\mathrm{nm}$.

---

[8] Assuming the convergence of the many-body state expansion for the given basis.

The expansion coefficients ξ_{ab} in Eq. (5.10) and Eq. (5.11) of all exciton states $|n,i\rangle$ $(i=1,2,3,4)$ are calculated by the Rayleigh-Ritz procedure, i. e., diagonalization of the exciton Hamiltonian $H^{(X)}$ over the entire cartesian basis:

$$\begin{aligned} H^{(X)} &= H_c - H_v - \frac{e^2}{4\pi\epsilon_0}\frac{1}{\epsilon_r(\mathbf{r}_c,\mathbf{r}_v)}\frac{1}{|\mathbf{r}_c-\mathbf{r}_v|} \quad , && (5.13)\\ H^{(X)}\,(\xi_{ab})_{n,i} &= E_n^{(X)}\,(\xi_{ab})_{n,i} \quad , && (5.14)\end{aligned}$$

where H is the single-particle Hamiltonian from Eq. (4.14), ϵ_r the dielectric function of the heterostructure, $\mathbf{r}$ the position operator, and $E^{(X)}$ the excitonic recombination energy. The subscripts c and v denote that the respective operators (H, $\mathbf{r}$) act only on the CB or VB part, respectively, of the exciton wave function Eq. (5.10). The matrix elements of $H^{(X)}$ are

$$\begin{aligned} H^{(X)}_{ab,jk} &= \left[E_a^{(c)} - E_b^{(v)}\right]\delta_{aj}\delta_{bk} && (5.15)\\ &- \frac{e^2}{4\pi\epsilon_0}\left\langle \Psi_a^{(c)}\Psi_b^{(v)}\right|\frac{1}{\epsilon_r(\mathbf{r}_c,\mathbf{r}_v)|\mathbf{r}_c-\mathbf{r}_v|}\left|\Psi_j^{(c)}\Psi_k^{(v)}\right\rangle \quad , \end{aligned}$$

where $E_a^{(c)}$ ($E_b^{(v)}$) is the eigenvalue of H belonging to $\Psi_a^{(c)}$ ($\Psi_b^{(v)}$), i. e., the single-electron (-hole) energy. Taking care of the preceeding expense considerations it is not recommended to evaluate Eq. (5.15) as is, since the calculation of the sixfold integrals is tedious and does not permit a convenient treatment of the dielectric function ϵ_r. A better way to calculate $H^{(X)}$ is to use Green's formula and decompose Eq. (5.15) into a solution of the Poisson equation (using the nested iteration technique [180]) and a subsequent volume integration:[9]

$$\begin{aligned} \nabla\left[\epsilon_r(\mathbf{r}_c)\,\nabla U^{(c)}(\mathbf{r}_c)\right] &= \frac{e^2}{4\pi\epsilon_0}\Psi_a^{(c)\dagger}(\mathbf{r}_c)\Psi_j^{(c)}(\mathbf{r}_c) \quad , && (5.16)\\ H^{(X)}_{ab,jk} &= \left[E_a^{(c)} - E_b^{(v)}\right]\delta_{aj}\delta_{bk} + \left\langle \Psi_b^{(v)}\right| U^{(c)} \left|\Psi_k^{(v)}\right\rangle . && (5.17)\end{aligned}$$

The wave-function product on the right-hand side of Eq. (5.16) is a scalar product of two eight-dimensional complex vectors of the type Eq. (4.15). By Eq. (5.16) the spatial variation of the dielectric constant ϵ_r , leading to the image charge effect, is taken into account. The decomposition of Eq. (5.15) is, in practice, a prerequisite for the tractability of the CI approach since the numerical accuracy required from the matrix elements is high in order to produce physically realistic many-body eigenstates (ξ_{ab}). The plausibility of calculated state vectors (ξ_{ab}) can be judged *a posteriori*, e. g., by checking whether the symmetries of the states or their dipole transitions are compatible with those of the QD.

---

[9] In Eq. (5.16) and Eq. (5.17) the roles of the CB and VB parts can be interchanged, of course.

The correlation effect is expressed by the existence of off-diagonal matrix elements $H^{(X)}_{ab,jk}$: Since $H^{(X)}$ in Eq. (5.14) is not diagonal its eigenvectors $(\xi_{ab})_{n,i}$ generally mix different electron-hole configurations $|\,\Psi^{(c)}_a\,\Psi^{(v)}_b\,\rangle$ within each exciton wave function $\Psi^{(X)}_{n,i}$ in Eq. (5.10). However, in strongly confining dots most often one configuration $|\,\Psi^{(c)}_a\,\Psi^{(v)}_b\,\rangle$ is dominant in a degenerate exciton state $|n,i\rangle$. This is characteristic to weak correlation. Weakly correlated exciton states $|n\rangle$ can still be classified by the one dominant CB $(|\,\Psi^{(c)}_a\,\rangle)$ and VB $(|\,\Psi^{(v)}_b\,\rangle)$ contribution, considering the Coulomb force an *energetic* correction only, as visualized by Fig. 6.2. Such a classification is useful for the interpretation of spectroscopy experiments which cannot resolve the fine structure due to exchange effects, like those discussed in Section 6.3.

5.2.5 Excitonic finestructure

The momentum matrix elements associated with the antisymmetric exciton basis vectors used in the Hartree-Fock approximation can be calculated from the electron picture. In the final state $|\,b\,\rangle$, the "vacuum" state, both electrons occupy the VB while in the initial state $|\,a\,\rangle$, the exciton, one electron (e. g., the second) is excited to the CB:

$$\begin{aligned}
|\,a\,\rangle &= (|\,e_{1v}e_{2v}\,\rangle - |\,e_{2v}e_{1v}\,\rangle)/\sqrt{2} \qquad , \\
|\,b\,\rangle &= (|\,e_{1v}e_{2c}\,\rangle - |\,e_{2c}e_{1v}\,\rangle)/\sqrt{2} \qquad , \\
\langle\,a\,|\,\mathbf{p}\,|\,b\,\rangle &= \langle\,a\,|\,\mathbf{p}_1+\mathbf{p}_2\,|\,b\,\rangle \\
&= \frac{1}{2}\begin{pmatrix} \langle\,e_{1v}e_{2v}\,|\,\mathbf{p}_1+\mathbf{p}_2\,|\,e_{1v}e_{2c}\,\rangle \\ -\langle\,e_{1v}e_{2v}\,|\,\mathbf{p}_1+\mathbf{p}_2\,|\,e_{2c}e_{1v}\,\rangle \\ -\langle\,e_{2v}e_{1v}\,|\,\mathbf{p}_1+\mathbf{p}_2\,|\,e_{1v}e_{2c}\,\rangle \\ +\langle\,e_{2v}e_{1v}\,|\,\mathbf{p}_1+\mathbf{p}_2\,|\,e_{2c}e_{1v}\,\rangle \end{pmatrix} \\
&= (\langle\,e_{1v}e_{2v}\,|\,\mathbf{p}_2\,|\,e_{1v}e_{2c}\,\rangle + \langle\,e_{2v}e_{1v}\,|\,\mathbf{p}_1\,|\,e_{2c}e_{1v}\,\rangle)/2 \qquad (5.18) \\
&= (\langle\,e_{1v}\,|\,e_{1v}\,\rangle\langle\,e_{2v}\,|\,\mathbf{p}\,|\,e_{2c}\,\rangle + \langle\,e_{2v}\,|\,\mathbf{p}\,|\,e_{2c}\,\rangle\langle\,e_{1v}\,|\,e_{1v}\,\rangle)/2 \\
&= \langle\,e_{2v}\,|\,\mathbf{p}\,|\,e_{2c}\,\rangle \qquad . \qquad (5.19)
\end{aligned}$$

The subscripts to $\mathbf{p}$ denote the spatial coordinate which $\mathbf{p}$ refers to. The reduction steps leading to Eq. (5.18) and Eq. (5.19) rely on the orthogonality of e_{1v} , e_{2v} , and e_{2c} (see p. 67). Eq. (5.19) implies that the momentum matrix element calculation needs not to distinguish between the Hartree and Hartree-Fock approximations for the exciton.[10] Hence, Eq. (5.11) holds also in the Hartree-Fock

---

[10] This is a particular case of a general theorem for matrix elements of sums of one-particle operators between determinantal states [88].

basis state	wave function	momentum ME	ΔM
b_1	$\left(\lvert c_1v_1\rangle - \lvert v_1c_1\rangle\right)/\sqrt{2}$	$\mathbf{p}_{11}$	-1
b_2	$\left(\lvert c_1v_2\rangle - \lvert v_2c_1\rangle\right)/\sqrt{2}$	$\mathbf{p}_{12}$	2
b_3	$\left(\lvert c_2v_1\rangle - \lvert v_1c_2\rangle\right)/\sqrt{2}$	$\mathbf{p}_{21}$	-2
b_4	$\left(\lvert c_2v_2\rangle - \lvert v_2c_2\rangle\right)/\sqrt{2}$	$\mathbf{p}_{22}$	1

Table 5.2:

Basis vectors for the Hartree-Fock calculation of the exciton ground-state finestructure (neglecting correlation effects) together with the associated momentum matrix elements (MEs), as well as the changes in the total angular momentum M, defined in hypothetical absence of strain and band mixing in the 0D wave functions.

approximation:

$$\Psi^{(X)} = \sum_{a,b} \xi_{ab} \left[\Psi_a^{(c)}(\mathbf{r}_1)\Psi_b^{(v)}(\mathbf{r}_2) - \Psi_b^{(v)}(\mathbf{r}_1)\Psi_a^{(c)}(\mathbf{r}_2)\right]/\sqrt{2} \quad , \tag{5.20}$$

$$\mathbf{p}^{(X)} = \sum_{a,b} \xi_{ab}\, \mathbf{p}_{ab} \quad . \tag{5.21}$$

In the strong confinement regime where excitons are weakly correlated and the single-particle level separations are much larger than the exchange energies the excitonic finestructure can qualitatively be understood neglecting correlation effects. Thus the CI Hamiltonian is expanded into four antisymmetric basis states, see Table 5.2, constructed from the single electron (c_1, c_2) and hole (v_1, v_2) ground states. In (hypothetical) absence of strain and band-mixing effects the single-particle states would have the total angular momenta $-1/2$ (c_1), $1/2$ (c_2), $3/2$ (v_1), and $-3/2$ (v_2), assuming bound holes to possess projections to the bulk HH band only. Then $\mathbf{p}_{12} = \mathbf{p}_{21} = 0$ in Table 5.2 and the coefficients ξ_2 , ξ_3 of b_2 , b_3 in Eq. (5.20) yield no contribution to the momentum matrix element in Eq. (5.21). Thus any state $\Psi^{(X)}$ with $\xi_1 = \xi_4 = 0$ in Eq. (5.20) is a *dark exciton*.

The time reversal symmetry in absence of magnetic fields causes the following matrix elements to vanish according to Eq. (4.23):

$$\begin{aligned}
\langle\, c_1v_1 \,|\, \mathcal{C} \,|\, c_1v_2 \,\rangle &= 0 \\
\langle\, c_1v_1 \,|\, \mathcal{C} \,|\, c_2v_1 \,\rangle &= 0 \\
\langle\, c_1v_1 \,|\, \mathcal{C} \,|\, c_2v_2 \,\rangle &= 0 \\
\langle\, c_1v_2 \,|\, \mathcal{C} \,|\, c_2v_1 \,\rangle &= 0 \\
\langle\, c_1v_2 \,|\, \mathcal{C} \,|\, c_2v_2 \,\rangle &= 0 \\
\langle\, c_2v_1 \,|\, \mathcal{C} \,|\, c_2v_2 \,\rangle &= 0 \qquad \text{where}
\end{aligned}$$

$$\mathcal{C} \;=\; \frac{e^2}{4\pi\epsilon_0\,\epsilon_r\,[\mathbf{r}^{(1)},\mathbf{r}^{(2)}]\,|\mathbf{r}^{(1)}-\mathbf{r}^{(2)}|}$$

(the superscripts denote the spatial coordinate which **r** refers to). The Hamiltonian for the lowest four exciton states, neglecting correlation and magnetic fields, but including SO interaction in the bulk, as well as strain and band mixing in the QD, expanded into the CI basis $\{b_1\,,b_2\,,b_3\,,b_4\}$ is thus

$$H^{(X_0)} \;=\; (E-C)\,\mathbf{1}_4 \;+\; \begin{pmatrix} x_1 & z_h & z_e & y_1 \\ \overline{z_h} & x_2 & y_2 & -z_e \\ \overline{z_e} & \overline{y_2} & x_2 & -z_h \\ \overline{y_1} & -\overline{z_e} & -\overline{z_h} & x_1 \end{pmatrix} \tag{5.22}$$

with the matrix elements displayed in Table 5.3. $E-C$ is the exciton ground-state energy according to the Hartree approximation, and the 4×4 matrix describes the finestructure due to the long-range exchange effect. Fig. 5.1 shows the energy level structure according to Eq. (5.22) in a pyramidal InAs/GaAs QD with a base length of 17 nm, calculated on varying levels of simplification.

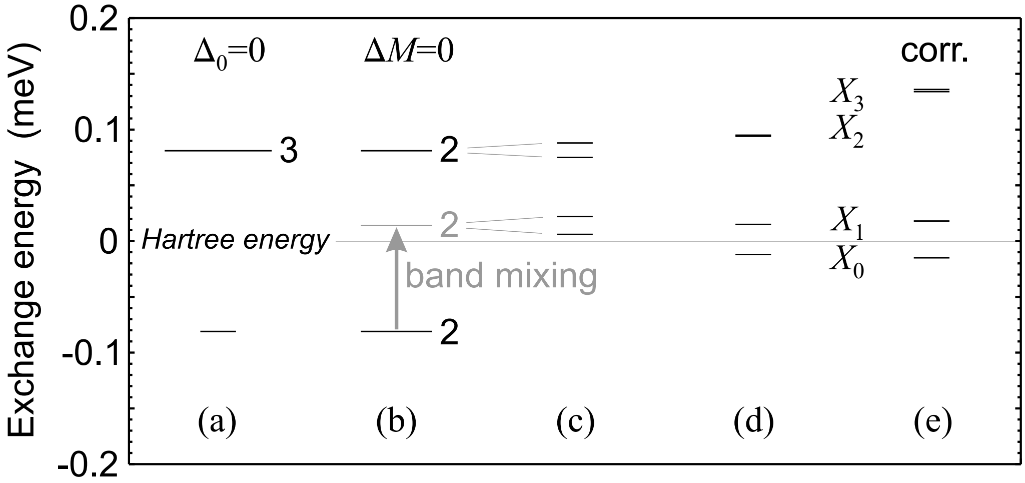


Figure 5.1:

Finestructure of the exciton ground state according to different models: (a) Neglection of the bulk SO interaction, $\Delta_0 = 0$; (b) $\Delta_0 \neq 0$, total angular momentum conservation in a HH exciton (black markers) and in presence of band mixing (gray marker and arrow); (c) inclusion of band mixing and simultaneous spin "flips" of electron and hole; (d) additional inclusion of individual spin "flips"; (e) additional inclusion of correlation. The integers and marker lengths indicate the multiplicities of the eigenvalues. The oscillator strengths of the four exciton states $X_0,\ldots,X_3$ according to (e) are shown in Fig. 5.2.

E	$E^{(c)} - E^{(v)}$	—	1042.6744
C	$\langle\, c_1v_1 \,\lvert\mathcal{C}\rvert\, c_1v_1 \,\rangle$, $\langle\, c_1v_2 \,\lvert\mathcal{C}\rvert\, c_1v_2 \,\rangle$, $\langle\, c_2v_1 \,\lvert\mathcal{C}\rvert\, c_2v_1 \,\rangle$, $\langle\, c_2v_2 \,\lvert\mathcal{C}\rvert\, c_2v_2 \,\rangle$	0	16.6200
C_e	$\langle\, c_1c_2 \,\lvert\mathcal{C}\rvert\, c_1c_2 \,\rangle$	0	22.5161
C_h	$\langle\, v_1v_2 \,\lvert\mathcal{C}\rvert\, v_1v_2 \,\rangle$	0	18.1906
x_1	$\langle\, c_1v_1 \,\lvert\mathcal{C}\rvert\, v_1c_1 \,\rangle$, $\langle\, c_2v_2 \,\lvert\mathcal{C}\rvert\, v_2c_2 \,\rangle$	0	0.0814
x_2	$\langle\, c_1v_2 \,\lvert\mathcal{C}\rvert\, v_2c_1 \,\rangle$, $\langle\, c_2v_1 \,\lvert\mathcal{C}\rvert\, v_1c_2 \,\rangle$	0	0.0141
y_1	$\langle\, c_1v_1 \,\lvert\mathcal{C}\rvert\, v_2c_2 \,\rangle$	±2	-0.0065 $-0.0003\,i$
y_2	$\langle\, c_1v_2 \,\lvert\mathcal{C}\rvert\, v_1c_2 \,\rangle$	±4	-0.0075 $-0.0028\,i$
z_e	$\langle\, c_1v_1 \,\lvert\mathcal{C}\rvert\, v_1c_2 \,\rangle$, $-\langle\, c_1v_2 \,\lvert\mathcal{C}\rvert\, v_2c_2 \,\rangle$	±1	0.0010 $-0.0089\,i$
z_h	$\langle\, c_1v_1 \,\lvert\mathcal{C}\rvert\, v_2c_1 \,\rangle$, $-\langle\, c_2v_1 \,\lvert\mathcal{C}\rvert\, v_2c_2 \,\rangle$	±2	-0.0001 $-0.0314\,i$

Table 5.3:

Definitions of the CI matrix elements in Eq. (5.22) and associated total angular momentum difference between the initial and final configurations (defined in hypothetical absence of strain and band mixing in the 0D wave functions). The numerical values (meV) refer to a pyramidal InAs/GaAs QD with a base length of 17 nm.

Neglection of SO interaction

In QDs formed from bulk materials without SO interaction (as considered in Ref. [37]), $\Delta_0 = 0$, the single-particle spin yields a good quantum number [89,181] and causes x_2 , y_1 , z_e , and z_h to vanish regardless of the dot shape, strain, or band mixing, while $x_1 = y_2$ due to the identical orbital parts of c_1 and c_2 (v_1 and v_2). Then Eq. (5.22) yields a spin singlet state with $E_0 = E - C - x_1$ and a spin triplet state with $E_1 = E - C + x_1$ and $2x_1$ is called the singlet-triplet splitting energy. The levels are shown in Fig. 5.1(a). As can be seen directly from the resulting matrix structure of $H^{(X_0)}$ the singlet state, being the ground state, and one of the triplet states are dark while the two other triplet states are bright. Hence only one exciton line at the triplet energy E_1 will appear in dipole spectra. The dark contribution to the triplet state cannot be isolated because of the level degeneracy.

As obvious from Table 5.3, $x_1 = y_2$ is a poor approximation to the actual situation governed by the SO interaction and band mixing effects.

Neglection of strain and band mixing

Including the SO interaction and assuming that the electrons have no VB projection ($c_1 = |1/2, -1/2\rangle$, $c_2 = |1/2, 1/2\rangle$) while the holes are pure HHs (having no CB projection, $v_1 = |3/2, 3/2\rangle$, $v_2 = |3/2, -3/2\rangle$) the long-range exchange matrix in Eq. (5.22) vanishes completely due to the total decoupling of the CB and VB. However, taking into account the CB-VB interaction perturbatively leads to the relation $x_1 = -x_2 \neq 0$ [182] reflecting the HH wave-function symmetry properties. Forcing, consistently, a conservation of the total angular momentum leads to the neglection of y_1 , y_2 , z_e , and z_h, as can be seen from Table 5.3, and yields two doublets at the energies $E_0 = E - C + x_2$ and $E_1 = E - C + x_1$. Since $x_2 = -x_1$ these doublets have the same energies as the previously discussed spin singlet and triplet states, respectively, as shown by the black markers in Fig. 5.1(b). The two-fold degenerate ground state is obviously dark, while the upper doublet is bright. Hence, the dark and bright excitons are energetically separated.

In the general case, the 0D wave functions mix all bulk bands, as detailed in Sections 3.2.1 and 4, so that $x_2 \neq -x_1$ (see Table 5.3). This affects mainly the energy of the dark excitons since the matrix element x_1 is absolutely largest and exists in any approximation for the exchange effect. The change of E_0 due to band mixing effects (still assuming M to be a good quantum number) is shown by the gray arrow and marker in Fig. 5.1(b) and visualizes the intrinsic limitation of approaches which employ perturbation theory on the basis of total angular momentum conservation [106,108,137,166,167,170,182]: Though the magnitudes of doublet *splittings* due to asymmetry effects may well be estimated, the energetic *positions* of the doublets with respect to the Hartree ground state are obscured by the neglection of band mixing.

Inclusion of band mixing

Both doublets are split into singlets by allowing simultaneous "flips" of the electron and hole spins, as represented by y_1 and y_2:

$$\begin{aligned}
E_0 &= E - C + x_2 - |y_2| \quad , \\
E_1 &= E - C + x_2 + |y_2| \quad , \\
E_2 &= E - C + x_1 - |y_1| \quad , \\
E_3 &= E - C + x_1 + |y_1| \quad ,
\end{aligned}$$

see Fig. 5.1(c). The energetic ordering of these levels is ensured by the relation $x_1 \gg x_2 > |y_1| \approx |y_2|$, see Table 5.3. Accordingly, $|\mathbf{p}_{11}| = |\mathbf{p}_{22}| \gg |\mathbf{p}_{12}| = |\mathbf{p}_{21}| > 0$, so that E_0 and E_1 still belong to "almost dark" excitons while E_2 and E_3 belong to bright excitons. The "dawn" of the dark exciton is only due to the partial relaxation of the selection rule $\mathbf{p}_{12} = \mathbf{p}_{21} = 0$.

A *mixing* of the dark and bright exciton states is accomplished by individual "flips" of the electron spin or the hole spin, represented by z_e and z_h, as can be seen from Eq. (5.22). The remarkable magnitude of z_h (see Table 5.3) for the studied QDs can be understood from the fact that the $\uparrow$ and $\downarrow$ spin projections of v_1 are 85% and 15%, while those of c_1 are 99.3% and 0.7%, respectively. Hence, the hole spin is much less well-defined in the QD than the electron spin. The individual spin "flip" transitions mainly push the two pairs of (almost) dark and bright states further apart, see Fig. 5.1(d), while mixing *all* momentum matrix elements in Eq. (5.21). As a secondary effect, the splitting of either level pair is modified. Still, the two lower exciton states yield much weaker dipole transitions than the two upper. The intensity ratio in the studied case is of the order $10^3 - 10^4$ and depends on the particular QD type. "Almost dark" exciton ground states resulting from the exchange effect have also been predicted for CdSe nanocrystals using a density-matrix approach in the frame of a tight-binding model [183].

Inclusion of correlation

To take into account correlation effects the approximation Eq. (5.22) is given up and a large basis CI calculation is performed using the 10 CB and 8 VB levels of the InAs pyramid with $b = 17\,\text{nm}$ shown in Fig. 4.8. The CI basis now contains 320 Slater determinants, instead of 4 in Eq. (5.22), and the calculation of the

51360 independent matrix elements of $H^{(X)}$ according to Eq. (5.17) requires (after economization) the evaluation of 666 wave-function products, 102720 volume integrations, and 530 solutions of the 3D complex Poisson equation Eq. (5.16).

Primarily, the direct Coulomb energy, denoted by C in Eq. (5.22), increases from 16.62 meV to 17.45 meV, but also the finestructure with respect to the ground-state energy in the Hartree approximation changes, as shown in Fig. 5.1(e). The resulting polarization dependent SMEs of the four exciton levels X_0, X_1, X_2, X_3 are shown in Fig. 5.2. For comparison, also the polarization dependencies in the (hypothetical) case of single-particle recombination, and for the exciton according to the Hartree approximation, are shown in Fig. 5.2.

The plots Fig. 5.2(a) and (b) differ by the effects of the Coulomb correlation which lowers the transition energy by 17.45 meV, reduces the in-plane polarization anisotropy[11] from 0.18 to 0.15, and increases the polarization-*averaged*[12] oscillator strength by 16% from 1.19 to 1.39. These Coulomb effects on the optical properties are relatively weak because of the strong confinement in the 17 nm large InAs/GaAs QDs. By inclusion of the Fermion character of the single particles (exchange effect) the formerly fourfold degenerate exciton splits up into the four levels shown in Fig. 5.1(e) having the different polarization characteristics shown in Fig. 5.2(c-f). The sum of their oscillator strengths – as calculated by analogy to Eq. (5.12) – yields a very similar polarization dependence and the same $\hat{\mathbf{e}}$ direction average as in the Hartree approximation shown in plot (b). Hence, the finestructure has no influence on the optical intensities in the presence of sufficiently strong broadening, but will be completely disguised then.

The proportions between the direct Coulomb interaction, correlation, and exchange effects on the energy structure of the exciton are visualized in Fig. 5.3. The largest contribution to the exciton ground-state energy, after (a) the sum of the strain induced band-edge shifts and the kinetic quantization (630 meV) plus the InAs band gap (413 meV), stems from (b) the direct Coulomb interaction (17 meV), followed by (c) the correlation (0.8 meV), and at least (d) the exchange effect, yielding a splitting of 0.15 meV between the almost dark ground state and the highest bright exciton in the quartet.

The calculated separation of 134 μeV between the average energies of the dark and bright exciton states agrees well with an experimental value of $\sim$ 200 μeV measured on self-assembled $\mathrm{In}_{0.4}\mathrm{Ga}_{0.6}\mathrm{As}$ dots with an average lateral size of $\sim$ 20 nm [106]. The fact that the (calculated) value for InAs QDs is smaller than the (measured) value for $\mathrm{In}_{0.4}\mathrm{Ga}_{0.6}\mathrm{As}$ dots is consistent with the material dependence of the bulk exchange splitting [184]. To conclude,

---

[11] Defined by Eq. (6.2).

[12] The average over all polarization directions is related to the "overall" oscillator strength relevant for spontaneous emission life times.

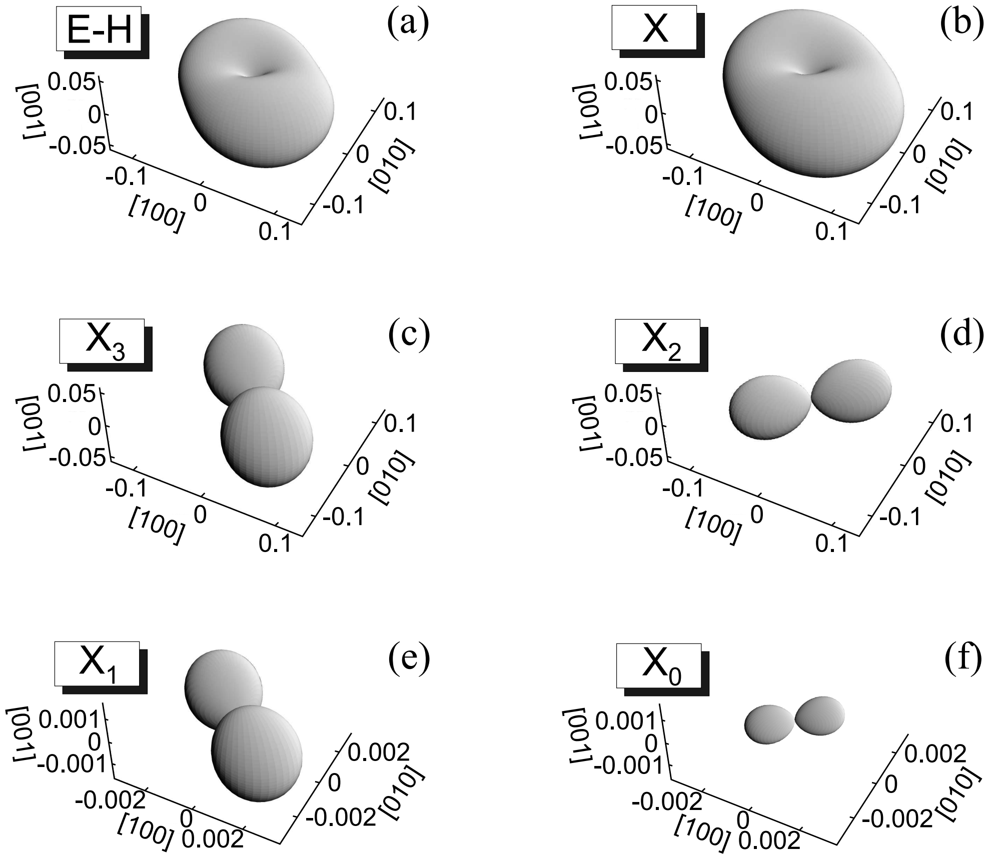


Figure 5.2:

SME according to Eq. (4.21) as function of the E-field direction $\hat{e}$ of the interacting light: (a) hypothetical single-particle recombination, calculated using Eq. (4.20) and Eq. (4.24); (b) excitonic recombination according to the Hartree approximation, calculated using Eq. (5.11) and Eq. (5.12); (c-f) excitonic recombination according to the Hartree-Fock approximation (including band mixing and correlation), calculated using Eq. (5.21) and shown in descending order of transition energies (the latter are shown in Fig. 5.1(e)). (c,d) belong to the almost degenerate bright exciton levels and (e,f) to the "almost dark" excitons (note the scale magnification by a factor 40).

- The 0D CB and VB single-particle ground states form four non-degenerate exciton states even in absence of magnetic fields, as a consequence of band-mixing effects in the confined wave functions.

- A non-degeneracy of exciton levels does not necessarily imply a structural asymmetry of the QD, although structural asymmetry lifts level degeneracies in case they exist in a corresponding, structurally symmetric system [106, 170, 182].

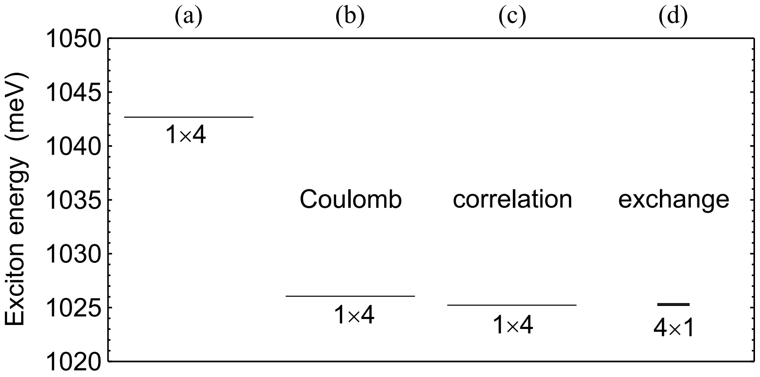


Figure 5.3:

Energies of the four lowest exciton states according to different models: (a) hypothetical single-particle recombination; (b) inclusion of direct Coulomb interaction, i. e., excitonic recombination neglecting exchange and correlation effects; (c) excitonic recombination neglecting exchange effects; (d) full calculation. The integers and marker lengths indicate the multiplicities of the eigenvalues. The finestructure is not resolved by this plot, except by the lifted level degeneracy.

- The finestructure splitting in presence of band mixing and correlation is significantly different from that predicted by perturbation theory proceeding from total angular-momentum conservation.
- The notion of bright and dark excitons remains valid at inclusion of strain, band mixing, and correlation, while the particular intensity ratios depend on the QD system.
- The finestructure is unlikely to be resolved in spectroscopy experiments on QD ensembles. A treatment of excitons in QD ensembles within the Hartree approximation appears sufficiently accurate.

5.2.6 Trion ground states

The finestructure collapses if the exciton is charged with one excess electron or hole, forming a negative (T^-) or positive (T^+) trion, respectively. The energy levels of trions are two-fold degenerate [109] since the individual spin "flip" energies of the two electrons (holes) in a negative (positive) trion cancel. Neglecting correlation, the ground-state recombination energies are

$$\begin{aligned} E^{(T^-)} &\approx E + C_e - 2C + x_1 + x_2 \qquad \text{and} \\ E^{(T^+)} &\approx E + C_h - 2C + x_1 + x_2 \qquad , \end{aligned}$$

using the notations from Table 5.3, Eq. (5.3), Eq. (5.4) and Eq. (5.5). Regarding the negative trion, the neglection of correlation causes an overestimation of $E^{(T^-)}$ by 1.9 meV for a 17 nm InAs pyramid, as compared to a large basis CI calculation. In contrast, the recombination energy $E^{(T^+)}$ of the positive trion is overestimated by 6 meV, like in the Hartree-LDA calculation (see p. 66). This underlines that the exchange-correlation effect of the holes is more pronounced than that of electrons in the studied example. Consequently, the positive trion in the 17 nm InAs pyramid is binding as compared to the exciton, yielding a spectral line on the low-energy side of the bright-exciton line, separated by 3.9 meV (the Hartree LDA predicts an anti-binding positive trion). Differently, the negative trion in the 17 nm InAs pyramid is anti-binding as compared to the exciton, since the large, positive difference between the direct Coulomb energies of both states, $C_e - C$, is not compensated by electron-electron exchange-correlation: The electron wave functions are more "stiff" than the hole ones, mainly due to their lighter mass. Hence, the negative trion yields a spectral line at a 4.6 meV higher energy than the exciton.

Since the lack of a continuum prevents multi-particle complexes like trions and biexcitons from dissociation, "anti-binding" does not imply instability in a QD: The confinement forces the formation also of energetically unfavourable states.

Generally, the trion energies depend strongly on the QD size [169], shape, and chemical composition [164], and this sensitivity is further enhanced in case of the biexciton. Both binding and anti-binding states are possible. The balance between the direct Coulomb interaction and the exchange-correlation effect requires an accurate calculation: The example of the positive trion shows that even a self-consistent LDA treatment may fail to correctly estimate the correlation energies.

Spectroscopic studies on $\sim$ 12 nm large, self-assembled InAs QDs revealed a "mystery" spectral line 1.8 meV below the exciton line whose most likely origin was considered a trion [109]. According to the present calculations this was probably a positive trion, in opposition to previous trion line assignments for such QDs based on self-consistent Hartree-LDA calculations [151]. Experimental evidence for the existence of binding negative trions was obtained on lens-shaped GaAs/AlGaAs QDs [185].

5.2.7 Biexciton energy levels

Neglecting correlation, the biexciton ground-state energy is

$$E^{(XX)} \approx 2E + C_e + C_h - 4C + 2(x_1 + x_2) \qquad ,$$

using the notations from Table 5.3. This is, however, an insufficient approximation as it predicts the biexciton to be *always* anti-binding. In the 17 nm InAs pyramid the exchange-correlation effect lowers the biexciton ground-state energy by 7 meV. Fig. 5.4(a) shows the lowest biexciton levels in these QDs calculated by large basis CI (675 Slater determinants) with and without the exchange effect.

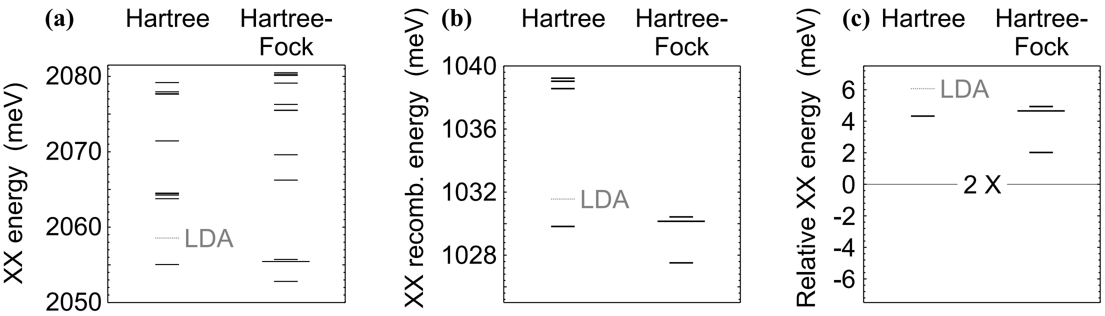


Figure 5.4:

(a) Lowest biexciton energy levels from large basis CI calculations in the Hartree and Hartree-Fock approximations, together with the ground state according to the Hartree LDA. (b) Recombination energies of the four lowest biexciton states. (c) Relative XX energy, i. e. $E^{(XX)} - 2E^{(X)} = -B_{XX}$. The first excited state is two-fold degenerate.

The Hartree approximation is fairly inaccurate, and the limitation of a separable ansatz for the biexciton wave function becomes clear from a comparison with the self-consistent LDA calculation from p. 65. In fact, the LDA yields a ground-state energy which is only 1.2 meV lower than by the above approximation neglecting correlation completely.

The ground state (Hartree-Fock approximation) consists to 80% of two electrons and holes in their single-particle ground states, the remaining 20% are intermixed excited-state configurations due to the correlation effect. The first two excited biexciton states shown in Fig. 5.4 are essentially due to the excitation of one hole from the ground state into the first excited state. The average excitation energy of the biexciton (2.7 meV) is comparable to the separation of the respective single-hole states (3.6 meV), but smaller due to the correlation effect. In the first excited biexciton state the ground-state hole and the excited hole have mainly parallel spin projections so that this level is two-fold degenerate. In the second excited, non-degenerate biexciton state the two holes have predominantly opposite spin projections.

Fig. 5.4(b) shows the predicted recombination energies $E^{(XX)} - E^{(X)}$ of the four lowest biexciton states. According to all used approximations the biexciton ground-state recombination occurs at higher energy than for the exciton, thus the biexciton is anti-binding in these dots, see Fig. 5.4(c), having a binding energy $2E^{(X)} - E^{(XX)}$ of -2 meV. Similar to the situation in the negative trion, the four-particle exchange-correlation is not sufficient to compensate the increase of the direct Coulomb energy, $C_e + C_h - 2C > 0$, when forcing two excitons into the QD.

The reason is that the correlation is restricted by the piezoelectric effect in the QD. The piezoelectric quadrupole potential shown in Fig. 4.7(d) separates electrons and holes, as seen in Fig. 4.11(a,b), thus simultaneously reducing the electron-hole attraction (and exchange) and increasing the pairwise Coulomb repulsion. In this ”piezoelectric regime” the biexciton is consequently anti-binding. Since the piezoelectric effect scales linearly with the QD size, the weaker charge separation in smaller pyramids may enable the exchange-correlation to overcome the net Coulomb repulsion, enabling the formation of binding biexcitons. The qualitative dependence of the biexciton binding energy on the QD size is illustrated in Fig. 5.5.

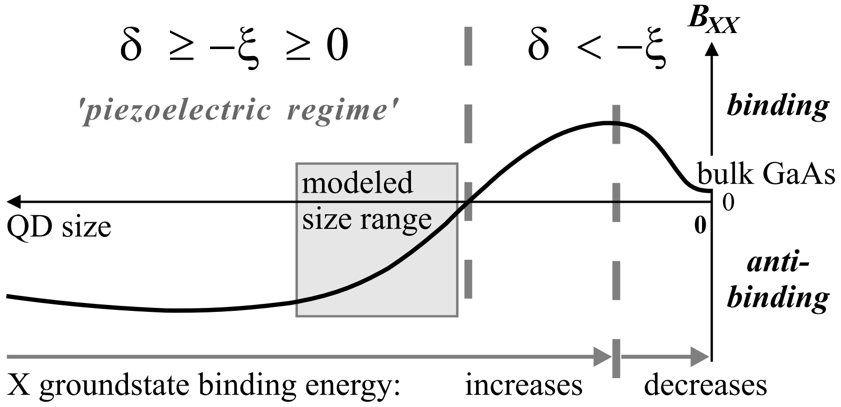


Figure 5.5:

Qualitative size dependence of the biexciton binding energy B_{XX} in piezoelectric, pyramidal InAs/GaAs QDs. δ denotes the direct Coulomb energy, ξ the sum of exchange and correlation energies.

The calculated size dependence of the biexciton binding energy is shown in Fig. 5.6 where the QD size is parametrized by the related exciton ground-state transition energy. Since the QDs with 10 nm base length are not eligible for a CI treatment they are not included. The break-even between the Coulomb and exchange-correlation effects is expected in InAs pyramid QDs with ~ 13 nm base length, having an exciton ground-state transition energy of ~ 1.1 eV at $T = 7$ K. For those dots, a resonance between the exciton and biexciton ground-state recombinations is predicted.

Indeed, weakly anti-binding biexcitons with a binding energy of -0.7 meV have been experimentally observed on ~ 12 nm large, self-assembled InAs/GaAs QDs

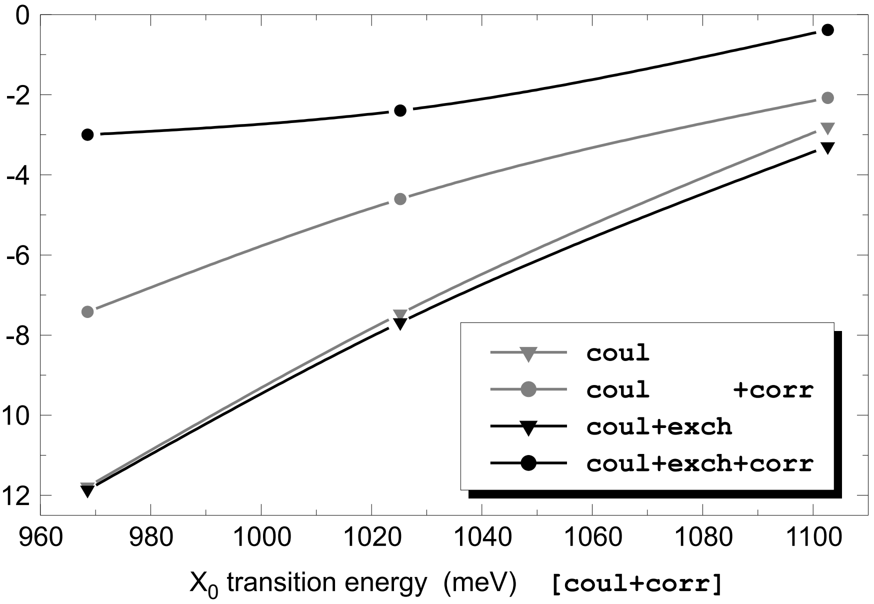


Figure 5.6:

Calculated size dependence of the biexciton ground state binding energy in piezoelectric, pyramidal InAs/GaAs QDs. The QD size is represented by the corresponding exciton ground state transition energy from the Hartree approximation. `coul` stands for the inclusion of direct Coulomb interaction, `exch` for the exchange effect, and `corr` for the inclusion of correlation effects.

[109]. Binding biexcitons with a binding energy of 3 meV have been observed on ∼ 20 nm large, self-assembled $In_{0.4}Ga_{0.6}As$/GaAs QDs [107]. This underlines that the effects of few-particle interactions depend sensitively on the structural, i. e., geometrical and chemical properties of the QD: In QDs which are less strained or have less developed facets, or consist of less piezoelectric materials, the C_{2v} (quadrupole) component of the confinement potential is weaker, favoring positive binding energies. A direct comparison between the present calculations and experiments [107,109,151] is limited by the deviation of the grown QDs from the InAs pyramid model used here.

Fig. 5.6 shows that the exchange effect has a negligible influence if correlation is excluded. In this case, the biexciton binding energies according to the Hartree and Hartree-Fock approximations differ only by approximately the singlet-triplet splitting energy $2x_1$. Under the impact of correlation this difference increases by one order of magnitude, demonstrating the coercitive interplay of exchange and correlation in the biexciton state.

To conclude, the biexciton binding energy is a highly sensitive, local probe for the confinement-dependent wave-function shapes in a QD.

Chapter 6

Optical Properties of Quantum Dots

The modeling of the optical properties of QDs and, moreover, self-organized QD ensembles is intricate even when restricted to the linear range. At sufficiently low temperatures the population of an individual QD with charge carriers is independent of that of others. No position-independent Fermi level exists and rate equations with average level populations are incorrect. In general, an ensemble of QDs has to be treated as a large set of microstates whose possible transitions are modeled by master equations [186, 187].

6.1 Calculated spectra

The absorption intensities and polarizations presented in this chapter are calculated using the CI method in the Hartree approximation as outlined in Section 5.2.4. In the spectroscopic experiments discussed in Section 6.3 unpolarized, vertically incident excitation was used, hence $\hat{e} \perp [001]$. The relevant SME is thus

$$\begin{aligned}
I &\propto \frac{1}{2\pi}\int_0^{2\pi} \left|\mathbf{p}_{[100]}\,\cos\phi + \mathbf{p}_{[010]}\,\sin\phi\right|^2 d\phi \\
&= \frac{1}{2}\left(|\mathbf{p}_{[100]}|^2 + |\mathbf{p}_{[010]}|^2\right) \\
&= |\mathbf{p}_{[100]}|^2 \qquad\qquad \text{where} \qquad\qquad (6.1)\\
\mathbf{p}_{[100]} &= (1,0,0)\cdot\mathbf{p}_n^{(X)} \quad , \\
\mathbf{p}_{[010]} &= (0,1,0)\cdot\mathbf{p}_n^{(X)} \quad ,
\end{aligned}$$

using $\mathbf{p}_n^{(X)}$ according to Eq. (5.12). I is proportional to the intensity for [100] polarized excitation due to the equivalence of the [100] and [010] directions in

(100) oriented pyramidal QDs.

6.1.1 Excitonic transitions

Fig. 6.1 shows the excitonic absorption spectra of QDs under vertical incidence of unpolarized light, dependent on the dot size. The transitions have been broadened artificially by Gaussians with 10 meV FWHM. Since, at least for the largest dots, not all single-hole levels were calculated the spectra for $b \geq 17.0\,\mathrm{nm}$ are incomplete at high energies.

The exciton spectra exhibit a large number of lines which mainly reflect the presence of several excited electron states. The assignment of the absorption lines to the respective single-particle transitions is also shown in Fig. 6.1 and justified by the weak correlation of excitons in the dots, as discussed in Section 5.2.4. Some electron states yield a significant oscillator strength with more than one hole state, thus contributing to more than one line. On the other hand, one line can be composed from several transitions, in particular, when the inhomogeneous broadening becomes $\geq 18\,\mathrm{meV}$, as it is in actual experiments.

A significant decrease of the SME of the ground-state transition with increasing dot size is found. This is due to the piezoelectrically induced symmetry breaking of the hole ground state reducing the overlap with the electron, see Fig. 4.11 (b) vs. (c). For large dots the transition becomes spatially indirect. This results in a decrease of the oscillator strength from 3.4 at $b = 10\,\mathrm{nm}$ down to 1.5 at $b = 20\,\mathrm{nm}$ (assuming vertically incident, unpolarized light).

Fig. 6.2 demonstrates that the Coulomb correlation introduces mainly an energetic correction to the single-particle spectra, underlining the weak correlation of exciton states in the strong confinement regime, as discussed on p. 78.

The general dependence of the SME on the E-field direction $\hat{e}$ according to Eq. (4.21) was shown in Fig. 5.2(b) for the ground-state transition in QDs with $b = 17\,\mathrm{nm}$. There is a strong anisotropy between the [001] and $[xy0]$ polarization directions, and a weaker anisotropy between [110] and $[1\bar{1}0]$. The latter reflects the C_{2v} symmetry of the single-particle wave functions and is somewhat reduced by the correlation of electron and hole in the exciton ground state, as seen by comparing Fig. 5.2(a) and (b). The excitonic transitions are generally expected to be less anisotropic than the (inexisting) recombination of "free" electrons and holes, due to the Coulomb correlation effect. The polarization anisotropy between the [110] and $[1\bar{1}0]$ directions,

$$p = \frac{I_{[1\bar{1}0]} - I_{[110]}}{I_{[1\bar{1}0]} + I_{[110]}} \quad , \tag{6.2}$$

of the ground-state transition is displayed in Table 5.1 for both cases, direct and

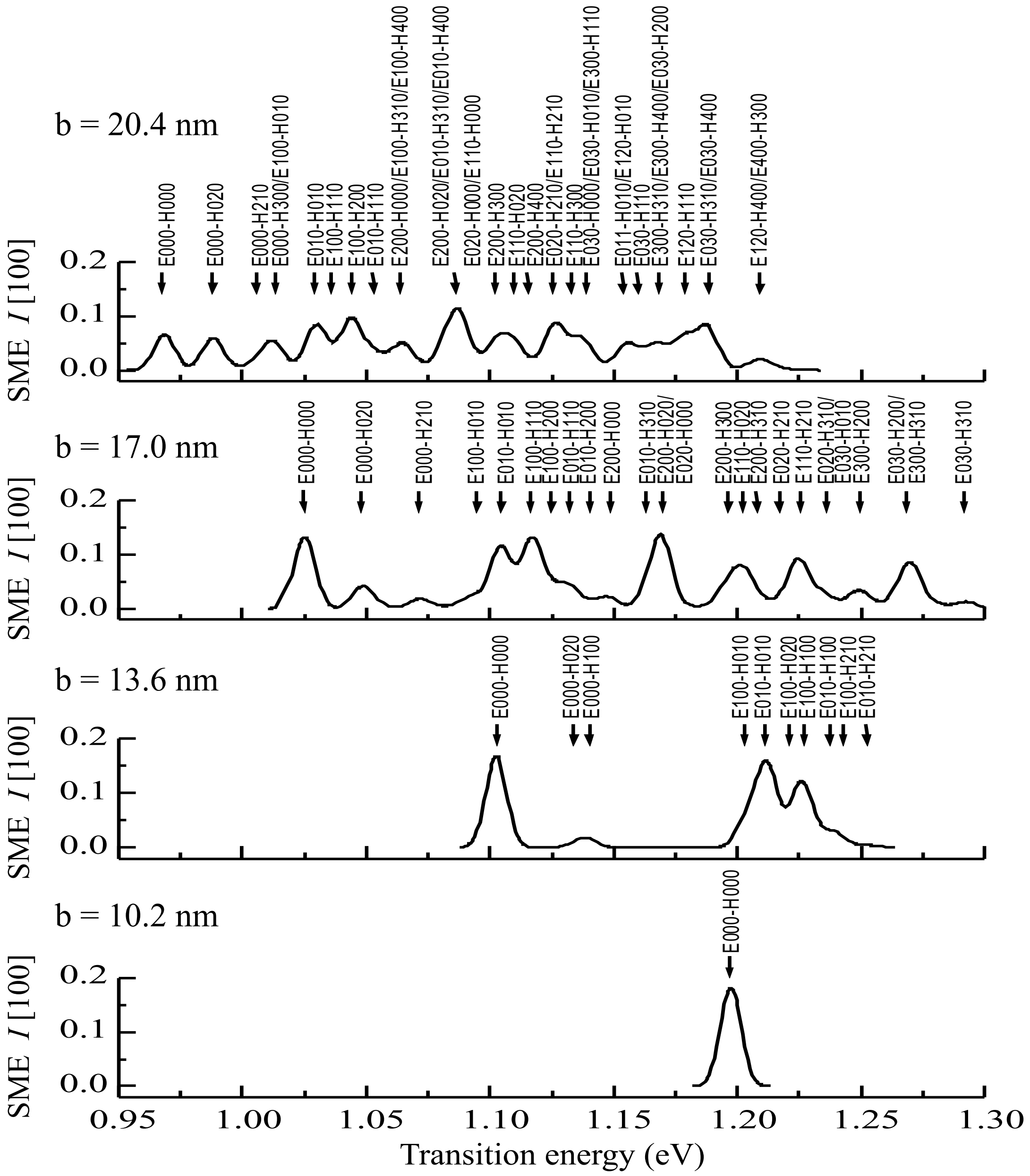


Figure 6.1:

Calculated excitonic absorption spectra of InAs pyramid QDs with different sizes b for unpolarized light under vertical incidence. The arrows with "quantum numbers" (see p. 50) indicate the strongest single-particle configurations contributing to the respective exciton. The SME I is calculated from Eq. (4.21) and Eq. (6.1). The absorption lines are artificial Gaussians with 10 meV FWHM. The seemingly missing transitions for $b = 20.4\,\text{nm}$ at energies above 1225 meV have not been calculated.

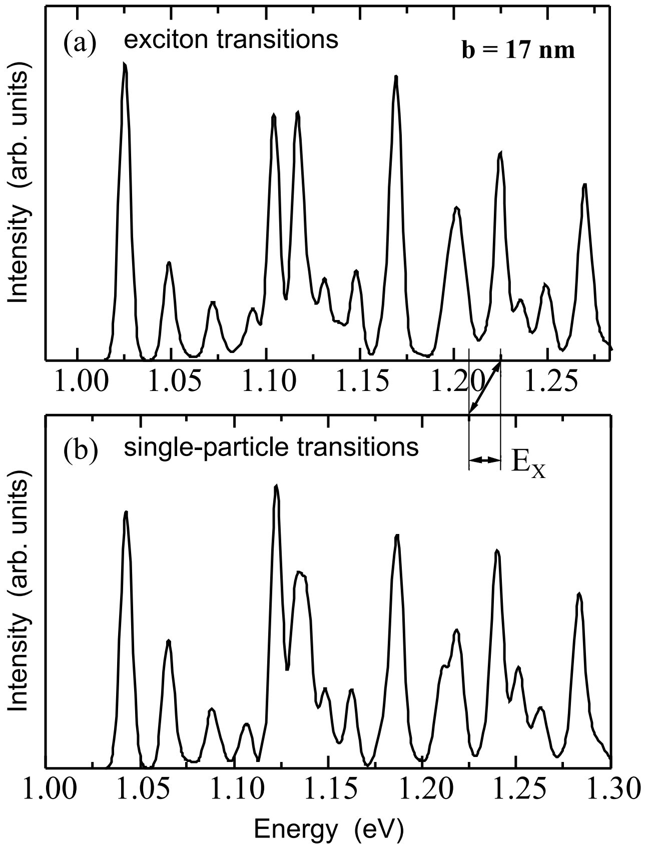


Figure 6.2:

Comparison of excitonic (a) and single-particle (b) unpolarized absorption spectra under vertical incidence for the QD size $b = 17\,\text{nm}$. The intensities are calculated from Eq. (4.21) and Eq. (4.24) or Eq. (5.12), respectively, and broadened artificially by Gaussians with 7 meV FWHM. The Coulomb correlation causes mainly an energetic shift of the transition lines while only modest intensity changes.

excitonic recombination, and as a function of the dot size b. p increases with b, and so does its reduction due to Coulomb correlation, in accordance with the weakening of the confinement in larger dots. The interaction with [001] polarized light is much weaker than under vertical incidence for all calculated transitions, as shown in Fig. 6.3.

6.1.2 Intraband transitions

The SMEs between electron levels are generally smaller than those representing exciton transitions, and yield a stronger anisotropy. Both properties are con-

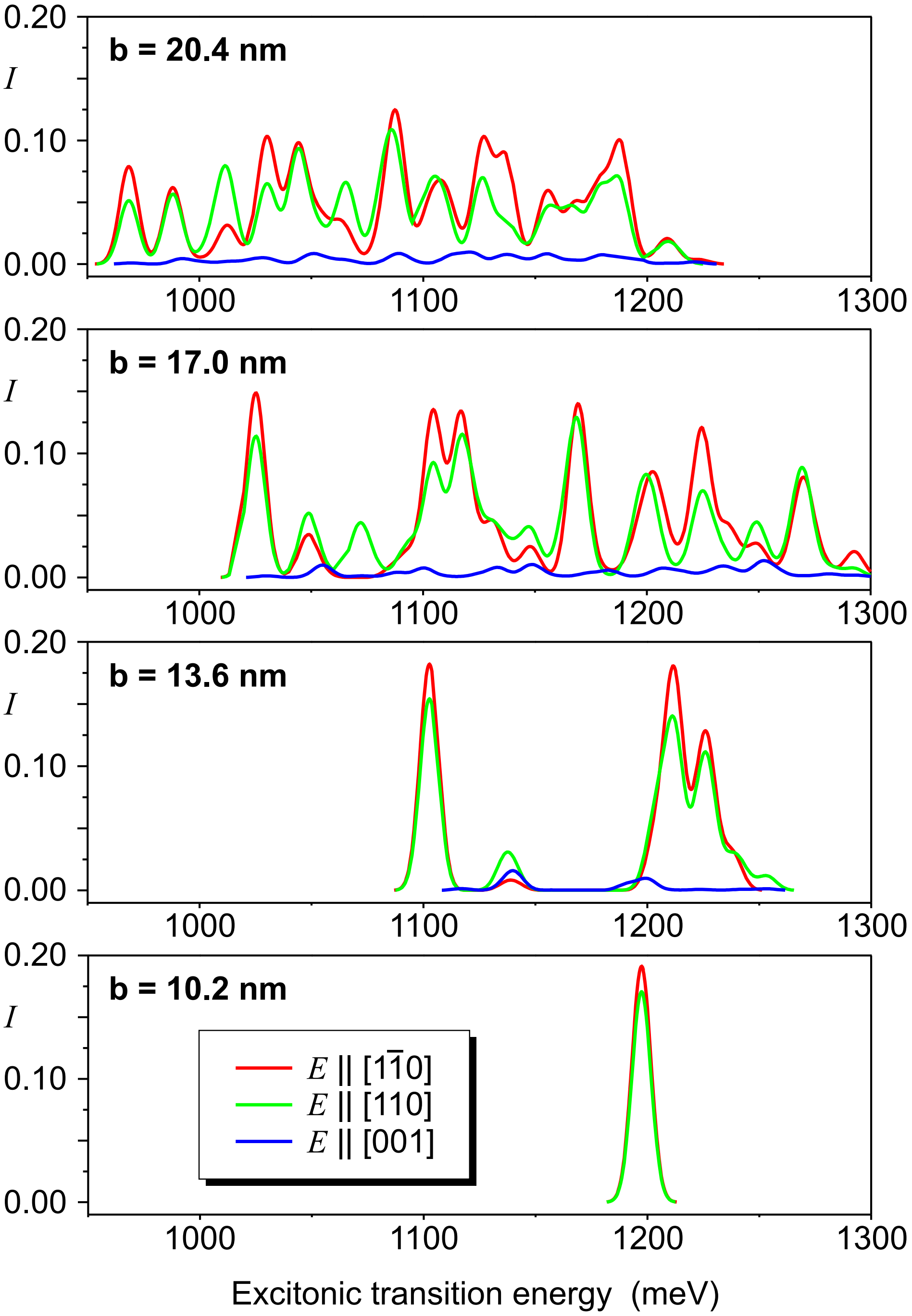


Figure 6.3:

Calculated excitonic absorption spectra for different QD sizes b and polarization directions. The SME I is given by Eq. (4.21) and Eq. (5.12). The absorption lines are artificial Gaussians with 10 meV FWHM. The seemingly missing transitions for $b = 20.4\,\text{nm}$ at energies above 1.2 eV have not been calculated.

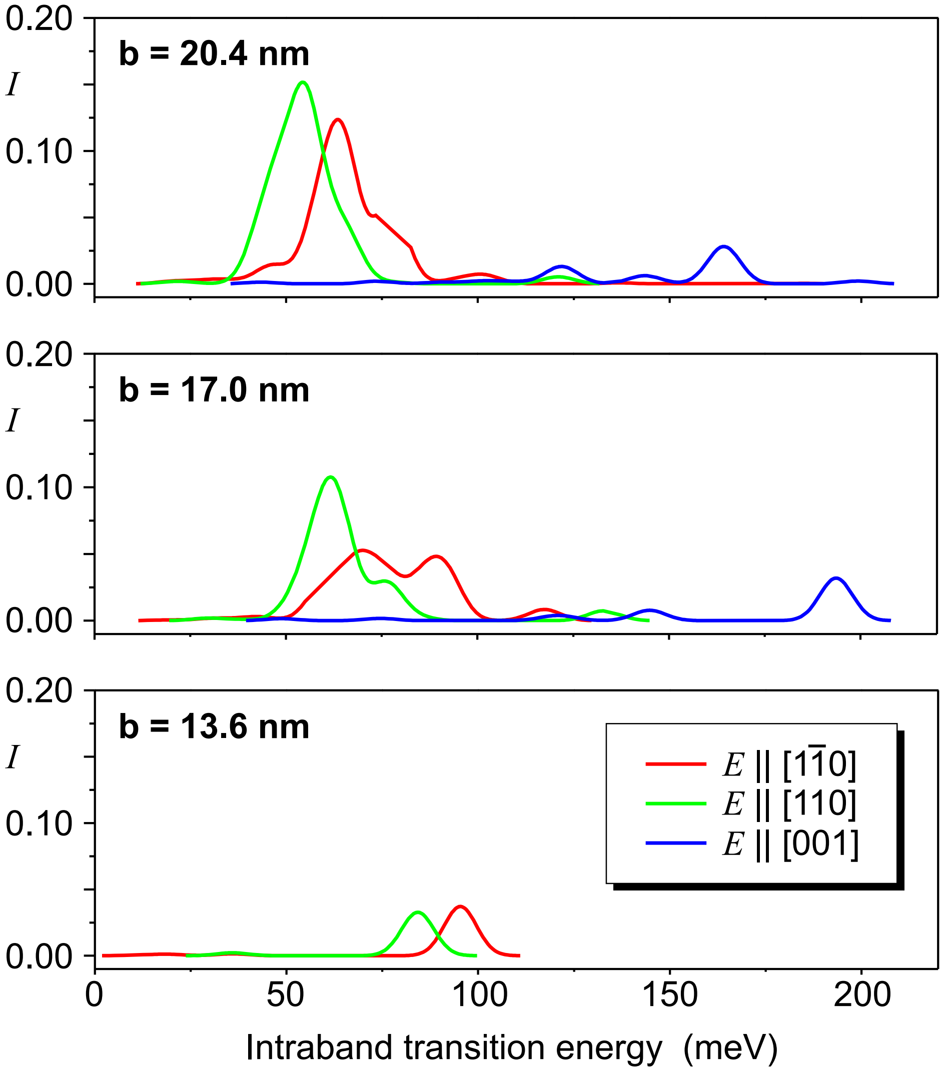


Figure 6.4:

Intraband absorption spectra, including electron-electron and hole-hole transitions, of InAs pyramid QDs of different sizes b and for different polarization directions. The SME I is calculated from Eq. (4.21), i. e. the explicit, reciprocal dependence of the absorption on the transition energy (see p. 58) is omitted in these plots. The absorption lines are artificial Gaussians with 10 meV FWHM.

sequences of the strict dominance of the s-type Bloch parts in the CB wave functions (see Fig. 4.10). The mid-infrared absorption spectra for different polarization directions are shown in Fig. 6.4 dependent on the dot size. The majority of excited electron states is involved in allowed mid-infrared optical transitions. The maximum of the mid-infrared absorption shifts to lower energies and becomes significantly stronger as the dot size increases, due to an increasing number of allowed transitions. This increase is additionally enhanced by the explicit energy dependence of the absorption (as discussed on p. 58). There is significant interaction with vertically incident light [188], which is a completely different situation from that in quantum wells. Thus mid-infrared light detectors for perpendicular incidence of light can be realized using QDs [189, 190].

6.2 Exciton phonon interaction

A first approximation to the coupling strength between the exciton ground state in the strong confinement regime and longitudinal-optical (LO) bulk phonons can be derived within an adiabatic treatment in the frame of the independent-phonon model [191]. Assuming the separation between the confined single-particle ground and first excited states to be larger than both the exciton binding energy and the bulk LO phonon energy, the independent-phonon model predicts (at zero temperature) a Poisson distribution of the excitonic oscillator strength over the phonon sidebands with the expectation value S called Huang-Rhys parameter [192]. The Huang-Rhys parameter for the inelastic LO-phonon scattering of the exciton ground state in the Hartree approximation is [193]

$$S = \frac{{f_0}^2}{(2\pi)^3(\hbar\,\omega_{\mathrm{LO}})^2}\iiint \frac{1}{|\mathbf{k}|^2}\left|\mathcal{F}\left(\sum_{a,b}\xi_{ab}\left[\overline{\sum_j \xi_{aj}\Psi_j^{(v)}}\,\Psi_b^{(v)} - \overline{\sum_i \xi_{ib}\Psi_i^{(c)}}\,\Psi_a^{(c)}\right]\right)\right|^2 d^3\mathbf{k}$$

using the exciton wave function Eq. (5.10), or

$$S \approx \frac{{f_0}^2}{(2\pi)^3(\hbar\,\omega_{\mathrm{LO}})^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{|\mathbf{k}|^2}\underbrace{\left|\mathcal{F}\left(|\Psi_e(\mathbf{r})|^2 - |\Psi_h(\mathbf{r})|^2\right)\right|^2}_{\Delta\psi(\mathbf{k})} d^3\mathbf{k} \qquad (6.3)$$

where $|\Psi_e\rangle$ and $|\Psi_h\rangle$ are self-consistent single-electron and -hole wave functions obtained from Eq. (5.1), $\mathcal{F}$ denotes the 3D Fourier transformation [180], $\hbar\,\omega_{\mathrm{LO}}$ is the bulk LO phonon energy of the QD material (InAs), and

$$f_0 = \sqrt{2\pi e^2\hbar\,\omega_{\mathrm{LO}}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right)}$$

[194] is the Fröhlich constant of the QD material, with ϵ_0 and ϵ_∞ being the static and high-frequency limits of the dielectric constant, respectively. The Bloch part of the wave functions is neglected in Eq. (6.3) in accordance with the basic envelope function approximation. Since the reference of Eq. (6.3) to the single-particle picture is based on the assumption of strong confinement widely suppressing the two-particle correlation, the separable ansatz underlying Eq. (5.1) to treat the Coulomb correlation seems appropriate.

As can be seen from Eq. (6.3) the Huang-Rhys parameter is sensitive to the local charge, i. e. the difference between the probability densities of the (self-consistent) electron and hole ground-states, which in the **k**-space is given by the underbraced term $\Delta\psi(\mathbf{k})$. The latter is significantly influenced by the piezoelectric effect reducing the spatial overlap of the two wave functions and thus the dipole component of their interaction with photons. In contrast, the interaction with LO phonons may be enhanced by the overlap reduction, due to the different density of states of the LO phonons: Since $\Delta\psi(\mathbf{k})$ is averaged by $1/|\mathbf{k}|^2$ the concentration of large $\Delta\psi$ values near the Γ point caused by the piezoelectric effect, see Fig. 6.5, results immediately in an increased Huang-Rhys parameter S. Analogously, the Coulomb correlation effect tendentially reduces S and is, hence, worth being included.

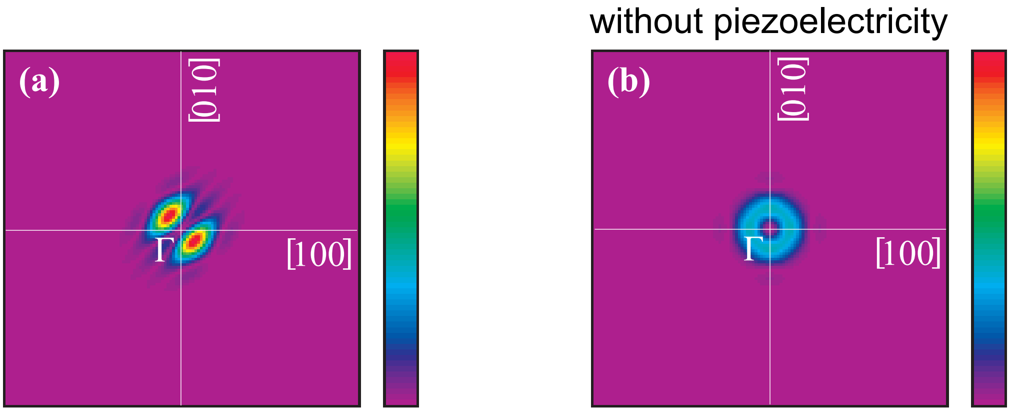


Figure 6.5:

Difference between the CB and VB ground-state probability densities, i. e. $\Delta\psi(\mathbf{k})$ from Eq. (6.3), plotted along the (001) plane of the **k** space, (a) including the piezoelectric effect and (b) neglecting it, calculated for the 17 nm large InAs pyramid QD. The shown **k**-space region is $7.3\,\mathrm{nm}^{-1} \times 7.3\,\mathrm{nm}^{-1}$ large. The color scales of both plots are identical, linear, and scaled to the maximum of plot (a).

To conclude, the polar exciton-LO-phonon interaction in piezoelectrically charged QDs is enhanced (as compared to the bulk case) [195–197] by the static quadrupole potential shown in Fig. 4.7(d). Since the piezoelectric effect is size dependent this enhancement is predicted to increase with the dot size, as shown in Fig. 6.6. The connection between the exciton ground-state transition energy and the Huang-Rhys parameter enables an experimental verification even without precise knowledge about the QD size. A weaker size dependence of S, caused by the strain distribution, is predicted neglecting piezoelectricity, see Fig. 6.6. As discussed in Ref. [195, 196] and shown in Fig. 6.6 the agreement between experiments (diamond markers) and the theory presented here is encouraging.

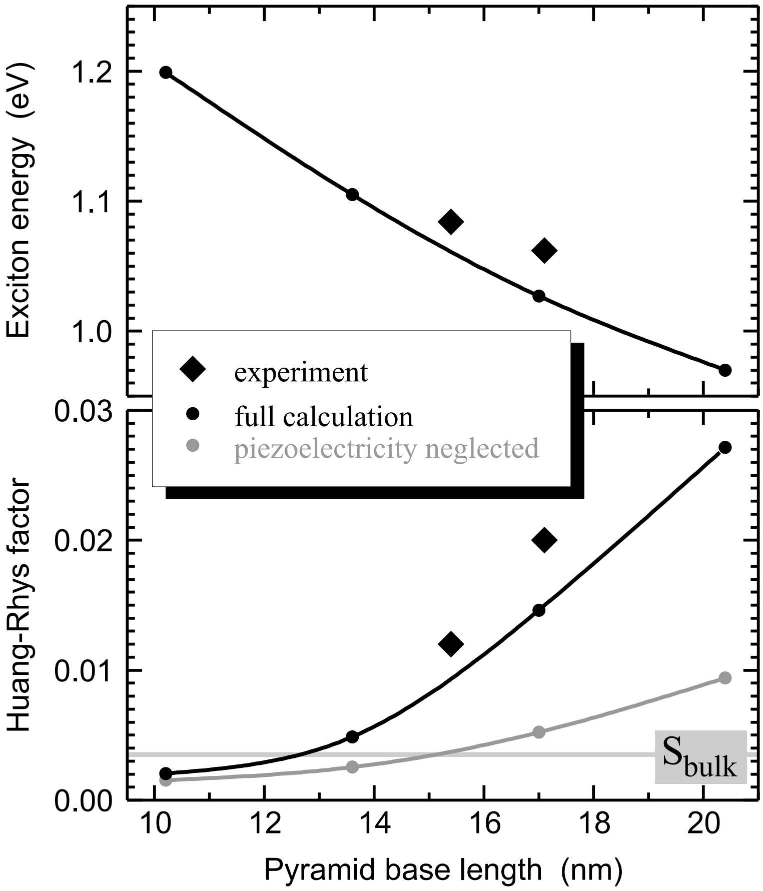


Figure 6.6:

Size dependencies of the exciton ground-state transition energy (top) and the Huang-Rhys parameter (bottom) in pyramidal InAs/GaAs QDs, calculated including the piezoelectric effect or neglecting it, together with experimental values [195, 196].

6.3 Comparison with experiments

A commonly taken experimental approach to the excited states in self-organized QDs is PL spectroscopy at high excitation densities, taking advantage of state filling in the discrete density of states [9, 10, 198–200]. However, such measurements probe the spectrum of highly populated QDs, which is subject to many-particle effects, and are additionally restricted by the inhomogeneous broadening of the QD ensemble. The experimental verification of detailed calculations requires (i) to overcome the inhomogeneous broadening in the optical spectra and (ii) to probe the absorption spectrum of empty QDs while (iii) knowing possibly much about the structural properties of the QDs.

The ensemble inhomogeneity is avoided in single-QD spectroscopy [91,106–109, 151,170,185,201–210], however, the comparison with calculations is hampered by the practically unknown structural properties of the actually investigated single QD. Alternatively, PLE spectroscopy on a QD ensemble allows to study the absorption of a subensemble of similar QDs defined by the detection energy. Such spectra are still inhomogeneously broadened due to shape and composition fluctuations but allow, at least, to study the dot-size dependence of excited exciton states, i. e. the quantum size effect, by tuning the detection window across the PL peak of the entire QD ensemble [211, 212]. Preconditions for a PLE study of the excited state spectrum of QDs are an excellent material quality and a high uniformity of the QDs. Attention must, in principle, be paid to the possibility that restricted intradot relaxation [213, 214] might influence the shape of the PLE spectra. However, multi-LO-phonon resonances can provide competing recombination processes which efficiently depopulate excited QD states [111,215].

In the following, near-resonant PLE spectra of self-organized InAs/GaAs samples with different average QD sizes [12] are compared to the calculated spectra shown in Fig. 6.1. Temperature-dependent measurements ensure that the PLE spectra do correspond to empty-dot absorption spectra, being suitable for studying the excited state transitions and their quantum size effect [12]. The spectroscopic experiments were performed by Robert Heitz on samples grown by Ilja Mukhametzanov and Anupam Madhukar and are described in detail in Ref. [12].

The various QD samples were grown by molecular beam epitaxy on semi-insulating GaAs(001) as described in Ref. [212, 216], depositing between 1.74 and 3.00 MLs InAs at 500°C. In order to vary the structural properties of the QD ensembles the InAs was deposited either continuously (C) [212] or in the variable deposition amount (V) approach using a 1.74 ML InAs QD seed layer [216]. The latter method produces vertical pairs of differently sized QDs in which only the larger QDs in the second layer are optically active [216, 217]. The samples are dubbed C and V, respectively, followed by the amount of InAs deposited for the active layer given in % of a ML (e. g., V300). The lateral size of the optically active QDs is $b \leq 19\,\mathrm{nm}$ and their height ranges from 4 to 10 nm, whereby the ensemble uniformity and the average aspect ratio of the islands increase along with the average QD size [12].

Fig. 6.7 compares the results of PLE measurements on the sample V300 with the calculations from Fig. 6.1. The excitation beam incidence (with a photon energy E_{exc}) on the sample is nearly perpendicular, i. e., close to the [001] direction and almost unpolarized, as assumed in Eq. (6.1). The detected exciton ground-state luminescence intensity at 7 K is plotted on a logarithmic scale in a contour plot as a function of the detection energy E_{det} and the excess excitation energy $\Delta E = E_{\mathrm{exc}} - E_{\mathrm{det}}$, and assumed to be proportional to the photon absorption due to exciton generation at E_{exc}.

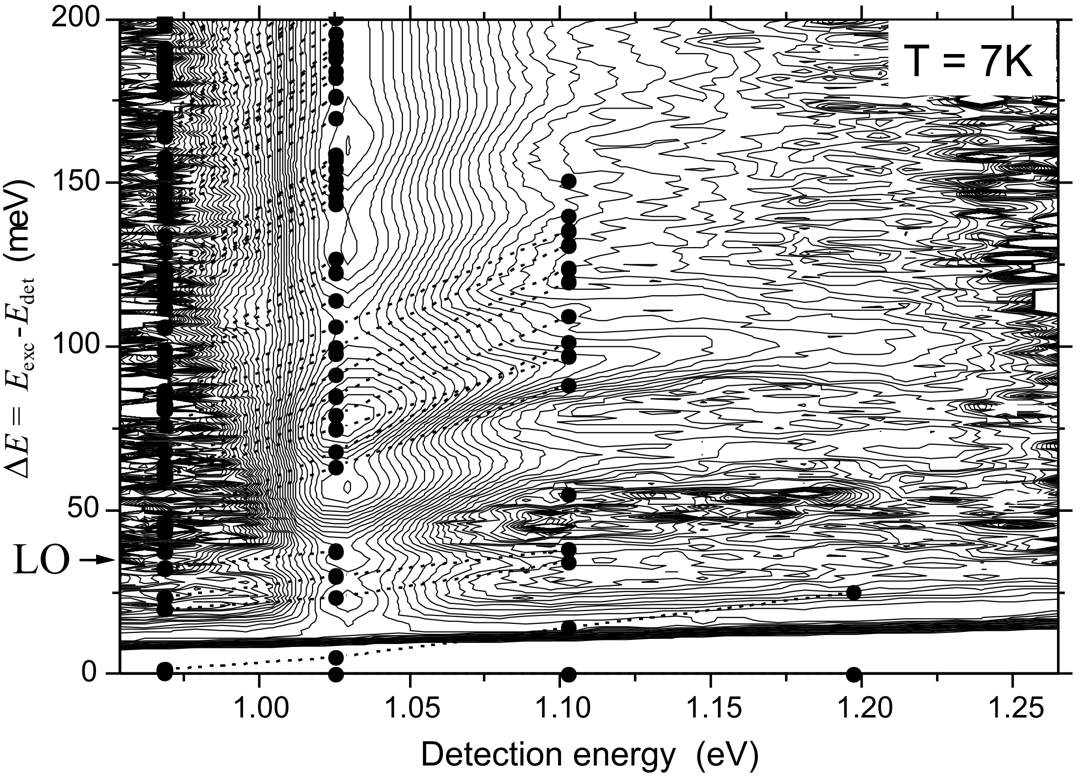


Figure 6.7:

PLE intensity of sample V300 at 7 K as contour plot on a logarithmic scale as function of the detection (E_{det}) and excess excitation ($\Delta E = E_{exc} - E_{det}$) energies. The detection energy E_{det} identifies the exciton ground-state transition energy and $E_{exc} = \Delta E + E_{det}$ is the energy of excited exciton states absorbing the excitation beam photons at a rate assumed to be proportional to the detected ground-state luminescence intensity. The arrow denotes the LO-phonon energy, i. e. the resonances from phonon-generated excitons in the ground state.

The quantum size effect is clearly observable: An increasing detection energy, i. e. ground-state transition energy, implies a decrease of the dot size and thus an enhanced separation between the confined exciton states, similar to the size dependence of the single-particle levels shown in Fig. 4.8. The black disks indicate the calculated exciton states. The dominant excitation resonance, observed at excess energies ΔE between 65 and 90 meV, is most intense in the energy gap between the two- and three-LO-phonon processes [215] which supports the identification of the excitation resonances with excited state transitions, rather than with phonon replica (the excess energy corresponding to the first LO-phonon sideband is shown by the arrow).

Fig. 6.8 compiles the excess excitation energies observed for the dominant PLE resonances in the two samples containing the largest QDs (C300 and V300). The most intense excitation resonances appear around excess energies ΔE of ~ 80 and $\sim 160\,\mathrm{meV}$, roughly corresponding to excited state transitions observed in inhomogeneously broadened high-density PL. Additional weaker resonances are resolved at ~ 22, ~ 60, ~ 110, ~ 130, and $\sim 190\,\mathrm{meV}$. Line shape fits reveal additional substructures for some of the resonances which have been omitted in Fig. 6.8 for the sake of clarity. The excited state transition spectrum of the QDs is considerably denser than expected from inhomogeneously broadened PL spectra.

The size-selective PLE spectra demonstrate the quantum size effect for large self-organized InAs/GaAs QDs with ground state transition energies below $\sim$ 1.13eV. For smaller QDs ($E_{\mathrm{det}} > 1.13\,\mathrm{eV}$), as also contained in some other of the investigated samples, the excitation resonance energies are *less* dependent on the detection energy, in contrast to the theoretically predicted steeper relation seen from Fig. 4.8. Partly even a reversed size dependence, i. e. an increasing excess excitation energy at increasing QD size, is observed. Since previous studies suggest that the observation of a quantum-size dependence requires highly uniform QDs [211] the evolution of the excited state spectrum for $E_{\mathrm{det}} > 1.13\,\mathrm{eV}$ shown in Fig. 6.8 is likely to depend not only on the quantum *size* effect, but also on changes in the average island shape and composition [218]. Therefore, the discussion here is restricted to samples with nominally 3.00 ML InAs deposition, containing the largest QDs and showing a clear quantum size effect [12]. These large QDs appear to be well approximated by the idealized InAs pyramids with $\{101\}$ side facets.

Certainly, the shape and composition of islands formed in the Stranski-Krastanow growth mode are controversial topics, especially in the case of overgrown structures for which In segregation and interdiffusion may occur during the cap layer growth. The persistent lack of detailed information on the structural parameters of the QDs, as well as on the statistical distribution of these parameters in an ensemble, is still a barrier for critical comparisons between experiments and detailed calculations. Nevertheless, the PLE spectra demonstrate the necessary progress established by all-numerical calculations, shown in Fig. 6.1 and Fig. 6.3, over simplistic approaches like the harmonic oscillator model predicting equally spaced transition lines: The observed wealth of absorption lines cannot be labeled by one common quantum number for multiply degenerate CB and VB states. The transitions might, however, be considered forming *groups* related to the quantum numbers of CB states, see Fig. 6.9(a), since the level separations of confined electrons are up to ~ 3 times larger than of the holes, while the splitting of the electron levels $|100\rangle$ and $|010\rangle$ is smaller than the quantization energies, as discussed in Section 4.3.1.

As seen from Fig. 6.8, the present calculations are in qualitative agreement with

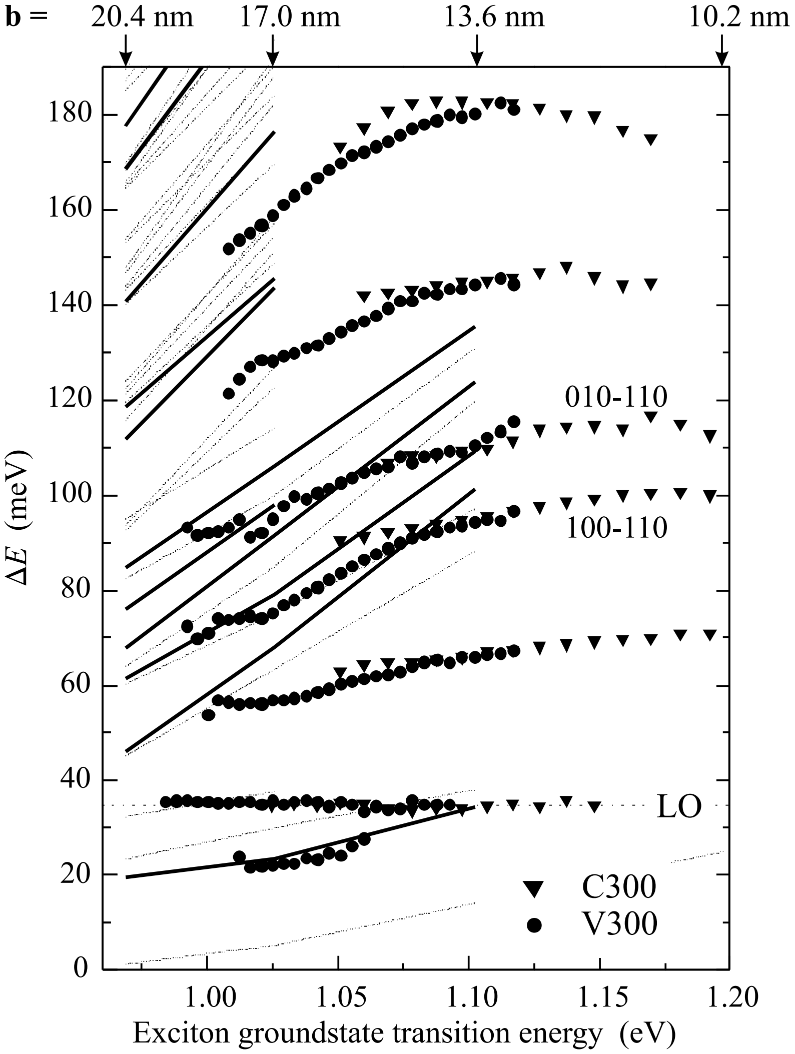


Figure 6.8:

Energetic positions of the dominant PLE resonances observed on the samples C300 (triangles) and V300 (disks). The lines represent the calculated exciton states, whereby solid lines indicate large SMEs (at least 10% of the maximal SME) and dotted lines correspond to weak transitions. The "quantum numbers" (see p. 50) refer to the dominant single-particle states ($e - h$) contributing to the respective resonances. The arrows atop of the plot indicate the pyramid base length b corresponding to the exciton ground-state energy according to the calculation. The first LO-phonon sideband is shown as a horizontal dotted line.

the PLE measurements regarding the larger islands ($E_{\mathrm{det}} \leq 1.13\,\mathrm{eV}$), predicting excited state absorption in those energy regions where excitation resonances are observed. However, the level separations for a given ground-state transition energy E_{det} are systematically overestimated. Accordingly, given a pyramid base length b the calculations predict a ground-state transition energy $30-40\,\mathrm{meV}$ below the experimental value [195], as already shown in Fig. 6.6. Hence, taking the experimentally known base length as reference – instead of the detection energy – introduces a phenomenological shift of the calculated ground-state transition energies by $\sim 40\,\mathrm{meV}$ to higher energies and provides for improved agreement with the experimentally observed excited state spectrum, as shown in Fig. 6.9.

Fig. 6.9(a) shows the calculated absorption spectrum for $b = 17\,\mathrm{nm}$. The agreement with the PLE spectrum detected at the predicted ground-state transition energy (1.025 eV) is only qualitative, see the solid line in Fig. 6.9(b). Better agreement is found for a PLE spectrum detected at 1.069 eV for which the experimentally estimated QD *size* ($b \approx 17\,\mathrm{nm}$ [195]) matches that of the calculated pyramid, see the dashed line in Fig. 6.9(b). A tentative assignment of the observed excitation resonances to excitonic transitions based on their energy and relative intensity is made in Ref. [12].

The resonances at $\Delta E = 50 \ldots 120\,\mathrm{meV}$ are attributed to transitions from the lowest excited CB states $|C100\rangle$ and $|C010\rangle$, being split by the C_{2v}-symmetry of the confinement (see Section 3.2.2), into different VB states. Fig. 6.10 displays the energy difference between the $|C100\rangle$-$|V110\rangle$ and $|C010\rangle$-$|V110\rangle$ exciton transitions as function of the ground state transition energy. This difference *increases* with increasing QD size, in contrast to the ordinary quantum size effect. Indeed, the reversed size-dependence is in accord with the theory, see Fig. 4.8, due to size dependence of the piezoelectric effect (see Section 4.2). The calculation (gray circles) predicts the correct slope but underestimates the magnitude. The ASA effect [38] discussed in Chapter 3 might account for the remaining discrepancy.

As already noted, the calculations fail to explain the PLE spectra of the smallest QDs in the samples ($E_{\mathrm{det}} > 1.13\,\mathrm{eV}$). The observed energy separations are significantly smaller than predicted[1] and seemingly lack any size dependence, see Fig. 6.8: The excitation resonances at large detection energies seem to assume a ladder corresponding to multiples of LO-phonon energies. Primarily, less structural information exists on the smaller islands, where shape and composition changes play a more important role than for large dots. For instance, interdiffusion decreases the confinement depth and increases the effective QD size, elevating the exciton ground state energy and reducing the excited state separation [219]. Second, enhanced interaction with non-radiative recombination centers in the barrier [215] or slower inter-level relaxation may occur in the smaller QDs.

The highest excited resonances show even a reversed size-dependence when the transition energy approaches that of the WL. Coulomb interaction between localized carriers in the QDs and their two-dimensional (2D) counterparts in the

---

[1] Part of this discrepancy may be related to the conventional choice of the $\mathbf{k}\cdot\mathbf{p}$ bandstructure parameters used in this part of the work (see Chapter 3 for discussion).

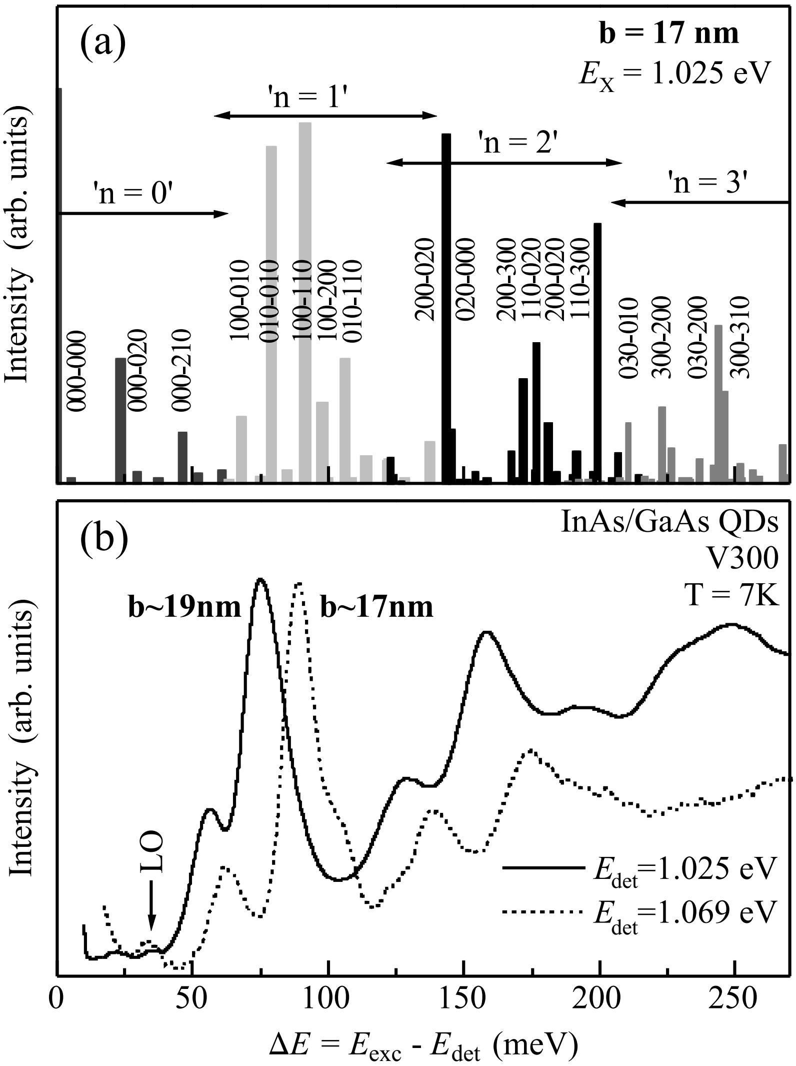


Figure 6.9:

(a) Calculated excitonic absorption spectrum for $b = 17\,\mathrm{nm}$, same data as in Fig. 6.1 and Fig. 6.2(a). The transitions are represented by bars with lengths indicating the relative magnitudes of the SMEs. The ”quantum numbers” (see p. 50) refer to the dominant single-particle configurations $(e - h)$ contributing to the respective excitons. The arrows group the transition lines according to the involved electron states, visualizing the prediction of quantum numbers n by a harmonic oscillator model for the electrons. (b) Low temperature PLE spectra of sample V300, either detected at the same exciton ground-state transition energy as in the calculation (solid line), or for the same dot size as in the calculation (dashed line).

WL would result in localized exciton states whereby the energy of the 2D carrier is – independently of the dot size – pinned to the WL, thus forcing a reversed

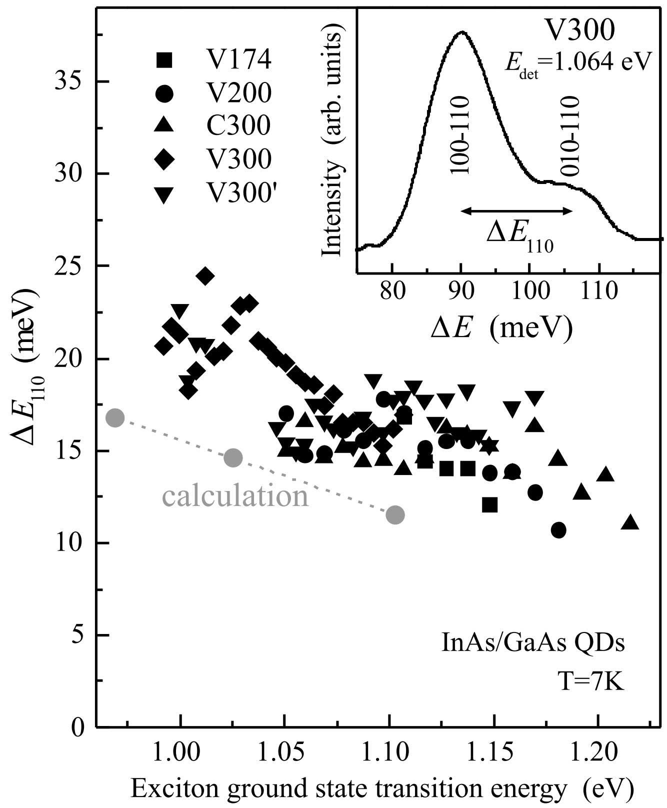


Figure 6.10:

Difference between the $|C100\rangle$-$|V110\rangle$ and $|C010\rangle$-$|V110\rangle$ exciton transition energies as function of the ground state transition energy, measured on various samples or calculated. The inset shows exemplarically how the difference was determined from the PLE spectra.

size dependence of the optical transition energy within a certain detection range.

In view of the uncertainties regarding the actual island shape and composition in the samples, as well as some of the material parameters entering the calculations (as discussed in Sections 2.3 and 4.3.2), the overall agreement is satisfactory. A comparably good agreement between **k·p** calculations and PLE measurements was reported for colloidal, spherical InAs nanocrystals [220]. Another, equally important test for the calculations is the splitting of the exciton localization energy between the CB and VB, i. e., the single-particle level positions with respect to the barrier band edges, as shown in Fig. 4.8. In optical spectroscopy of exciton transitions such information is only indirectly accessible, but it can be derived from capacitance spectroscopy. Investigations of electron and hole localization energies in self-organized InAs/GaAs QDs by deep level transient spectroscopy demonstrate good agreement with the present predictions [95, 221].

6.4 Resumé

The detailed theory of the electronic structure and optical properties of self-organized InAs/GaAs QDs outlined in this part is coroborated by several experiments. A variety of predictions has been confirmed semiquantitatively, whereby the inevitable discrepancies between measured quantities and predicted values are small enough to establish a validation of the theory:

- Single-particle level positions with respect to the barrier band edges, measured by deep level transient spectroscopy [95, 221],
- optical transitions from excited exciton states, measured by PLE [12],
- exchange splitting (separation between the average dark and bright exciton energies) and polarization properties of the exciton ground state, measured by single-QD magnetoluminescence [106],
- recombination energies of trions, measured by excitation dependent PL [109],
- sensitive dependence of the biexciton binding energy on the geometrical and chemical structure of the QD, evidenced by the observation of binding [107] and anti-binding [109] biexcitons,
- enhanced polar exciton-LO-phonon interaction, measured by PL [195].

The present treatment advances beyond previous work as it predicts all above mentioned properties within *one coherent frame work*, and based on *one and the same QD model*, rather than employing individually tailored models for each of the observed effects, thus establishing the mutual consistency of the various properties. In addition, *the theory does not lean on adjustable parameters*, although it is flexible to take into account arbitrary geometrical shapes and chemical compositions of QDs.

Its predictive power is limited by the lack of sufficiently detailed information on the geometrical and chemical structure of the QDs, and by the uncertainty regarding several bulk properties of the constituent materials (as detailed in Section 2.3, Chapter 3, and Section 4.3.2).

The eight-band $\mathbf{k}\cdot\mathbf{p}$ model adapted to QD calculations, in connection with different many-particle interaction models, is a valid and capable tool aside of experimental techniques for the investigation of epitaxial semiconductor QD heterostructures.

Chapter 7

Bandstructure in Quantum Wires

Semiconductor QWRs have also attracted interest for laser applications because of their potential for a reduction of the threshold current density [222] and the increased gain and differential gain achieved by 2D confinement [8]. Technological progress in the epitaxial fabrication of lattice-mismatched single and multiple QWRs on non-planar substrates led to various types of pseudomorphic QWRs among which V-groove QWRs have been intensively studied [6,7,13–16,223–228]. In the frame of this work, various representants of this QWR type were studied theoretically on the basis of the eight-band $\mathbf{k}\cdot\mathbf{p}$ model. Several aspects of single-particle confinement that have been discussed in Chapter 4 regarding the 3D confinement in QDs also apply to the 2D confinement in QWRs. The following considerations hence focus on the differences between QWRs and QDs originating from the existence of free, 1D carrier motion along the wire direction.

To briefly recapitulate the basic electronic properties of QWRs [53, 229] a straight wire of infinite length is considered, assuming the lateral confinement profile to be constant along the wire. Choosing the z-axis parallel to the wire the wave function of a charge carrier may be written as a product $\phi(z)\cdot\psi(x,y)$ where $\phi(z)$ describes the unconfined part of the wave function and $\psi(x,y)$ is the 2D confined part. Within the single-band effective-mass approximation using parabolic bands the confined part $\psi(x,y)$ of an electron wave function is the solution of the 2D "Schrödinger" equation $H_{\mathrm{2d}}\,\psi(x,y) = E_i\,\psi(x,y)$ with i being the quantum number and E_i the energy of the confined 1D state,

$$H_{\mathrm{2d}}(x,y) \;=\; -\frac{\hbar^2}{2}\nabla\left[\frac{1}{m^*(x,y)}\nabla\right] + V(x,y) \qquad ,$$

and m^* being the spatially dependent electron effective mass in the strained bulk material. V is the spatial bulk CB edge profile. The unconfined part ϕ of the wave function can be described by plane waves $\phi(z) = \frac{1}{\zeta}\exp\left(\mathrm{i}\,k_z\,z\right)$ with ζ being a scaling factor. Accordingly, the longitudinal motion of electrons is described by a 1D wave "vector" k_z yielding a parabolic longitudinal energy dispersion:

$$E^e_{\mathrm{QWR}}(k_z) \;=\; E_i + \frac{\hbar^2 k_z^2}{2m^*_{\mathrm{1D}}} \tag{7.1}$$

where m^*_{1D} is the 1D, longitudinal effective mass. Unlike the situation in QDs, each 2D quantized state i can be populated by more than two electrons, due to the k_z-continuous term in Eq. (7.1), so that electrons belonging to the same 1D state i actually form subbands. For $k_z \neq 0$ these are non-degenerate even at zero magnetic fields, in contrast to the 0D single-particle levels in QDs. The resulting density of states (DOS) is piecewise continuous:

$$\rho_{\mathrm{QWR}}(E) \;=\; \sqrt{\frac{2m^*_{\mathrm{1D}}}{\pi^2\hbar^2}} \sum_i \frac{1}{\sqrt{E-E_i}} \qquad .$$

In thermal equilibrium each subband has a thermal population with an effective carrier temperature T and, assuming infinitely high confining barriers V, the number of electrons in each subband i is

$$n^{(i)}_{\mathrm{QWR}} \;=\; \int_0^\infty \rho_{\mathrm{QWR}}(E) f(E)\,\mathrm{d}E \;=\; N^{(i)}_{\mathrm{QWR}}\, F_{-1/2}\left(\frac{E_F - E_i}{k_B T}\right) \qquad ,$$

where f is the Fermi distribution,

$$F_{-1/2}(z) \;=\; \int_0^\infty \frac{\mathrm{d}x}{\sqrt{x}\,[1+\exp(x-z)]}$$

is the Fermi integral of order $-\frac{1}{2}$, and

$$N^{(i)}_{\mathrm{QWR}} \;=\; \sqrt{\frac{2m^*_{\mathrm{1D}} k_B T}{\pi^2\hbar^2}}$$

is the effective 1D DOS of the i-th subband (k_B is the Boltzmann constant). Like in QDs, many-particle effects gain importance with increasing population of the 1D subbands, whereby the total Coulomb charge of the confined carriers may alter the lateral confinement potential V. The latter effect can be taken into account by self-consistent mean field calculations [128,129,230–232]. Furthermore, effects similar to the band-gap renormalization in bulk semiconductors have been observed [233] whose interpretation is controversial [234–237]. Caused by growth effects, the lateral shape of the QWR usually changes slightly along the wire and, in QWR arrays, from wire to wire, leading to inhomogeneous broadening of the level energies E_i. Additionally, longitudinal potential fluctuations ($\partial V/\partial z \neq 0$) may localize charge carriers at low temperatures so that the QWR electronically breaks up into an irregular chain of QDs [238].

7.1 Eight-band k·p model of strained quantum wires

In the following, the 1D subband structure as well as the CB, VB, and joint DOS and the confined wave-function parts ψ at the subband edges of V-shaped $\mathrm{In}_{0.2}\mathrm{Ga}_{0.8}\mathrm{As}/\mathrm{Al}_c\mathrm{Ga}_{1-c}\mathrm{As}$ QWRs are calculated by eight-band $\mathbf{k}\cdot\mathbf{p}$ theory, as reported in Ref. [40]. Like for strained QDs, a pronounced influence of the piezoelectric potential on the bandstructure is found, aside of a marked spin-splitting of the 1D VBs and non-parabolicity of the 1D CB.

The V-groove QWR with the nominal composition $\mathrm{In}_{0.2}\mathrm{Ga}_{0.8}\mathrm{As}$ is surrounded by a graded composition barrier of $\mathrm{Al}_c\mathrm{Ga}_{1-c}\mathrm{As}$ whose aluminium content c increases linearly with the vertical distance from the heterointerface, as visualized in Fig. 7.1(a). Within 200 nm from the interface, c increases from 0.2 to 0.7, i. e., the QWR and barrier form a "graded index separate confinement heterostructure" (GRINSCH). The cross section shape of the wire is taken directly from a transmission electron microscopy image and digitized with 1 nm resolution, yielding 13563 pixels. The growth direction [001] is assigned to the geometrical coordinate y. The wire is elongated along the $[1\bar{1}0]$ direction assigned to the coordinate z, while x refers to [110].

The lattice mismatch of +1.4% leads to mainly compressive strain in the QWR which is much weaker than the strain in InAs/GaAs QDs. The strain distribution shown in Fig. 7.1(c-f) is calculated within the CM model, yielding the strain tensor $\hat{\epsilon}$ with respect to the principal crystal axes as a function of the position (x, y) in the QWR cross section plane $(1\bar{1}0)$ [239]. From $\hat{\epsilon}(x, y)$ the piezoelectric potential $V_\mathrm{p}(x, y)$ shown in Fig. 7.1(b) is calculated, possessing the same translation symmetry as the QWR and, hence, being independent of z.

For the calculation of the continuous 1D-bandstructure in the direction z, where carrier transport is possible, wave vector components k_x, k_y, k_z with respect to the geometrical coordinates are introduced,[1] in order to take advantage of the translational symmetry along $[1\bar{1}0]$. Thus the envelope functions are separable, as outlined above, into a plane wave $\phi(z)$ with the wave number k_z and a confined part $\Psi(x, y)$ defined in the cross section plane. In the spatial representation of the Hamiltonian, $k_x \mapsto -\mathrm{i}\,\partial_x$ and $k_y \mapsto -\mathrm{i}\,\partial_y$ while k_z remains a real number. Thus a system

$$\hat{H}\left(x, y, k_z, \partial_x, \partial_y, \partial_{xx}, \partial_{xy}, \partial_{yy}\right)\Psi(x, y) \;=\; E\,\Psi(x, y) \tag{7.2}$$

of eight partial differential equations for the eight envelope functions $\psi_{s\uparrow}$, $\psi_{x\uparrow}$,

---

[1] k_x, k_y, and k_z are related to the principal crystal axes by the same rotation matrix as the chosen local coordinate system (x, y, z).

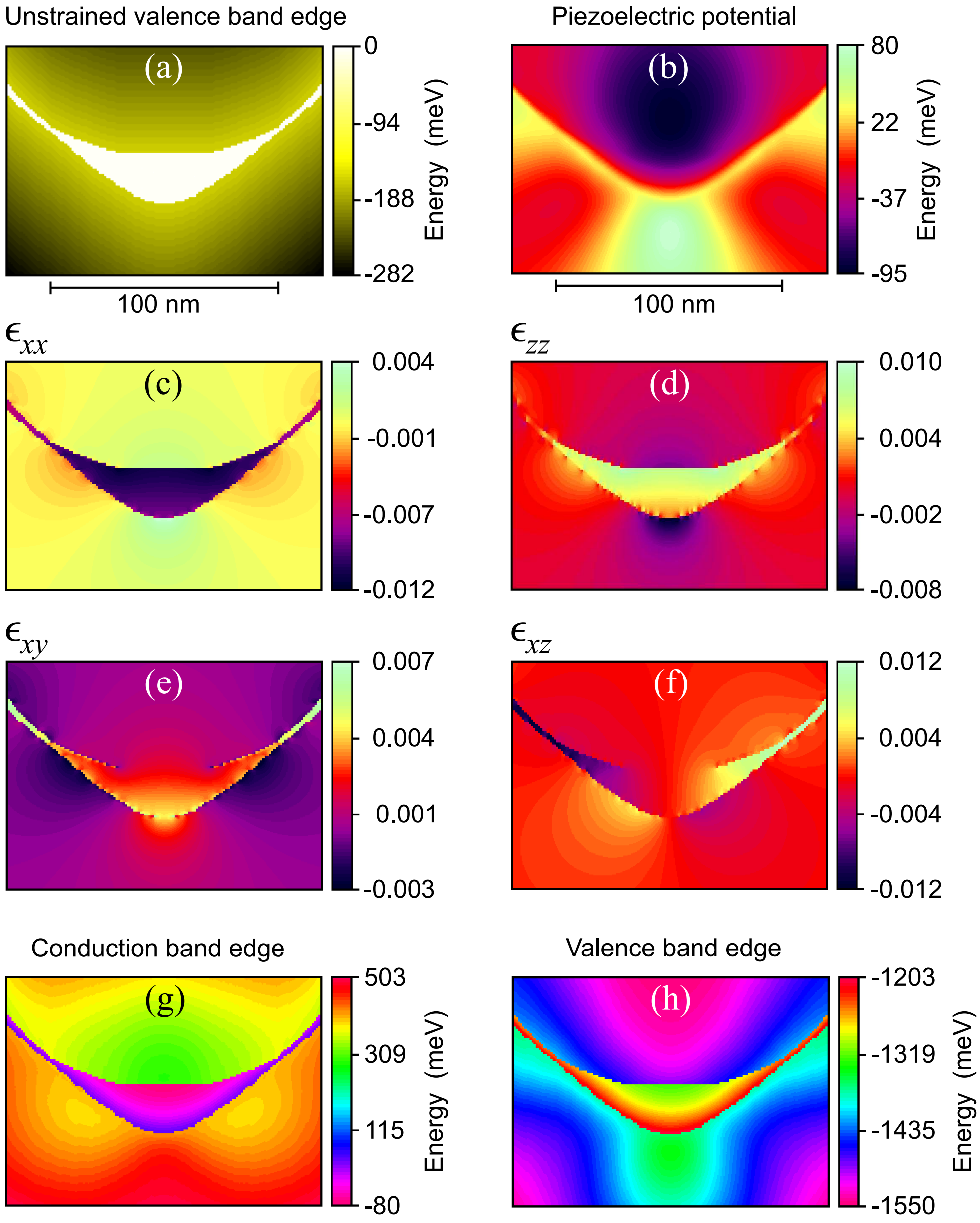


Figure 7.1:

(a) The unstrained bulk VB edge profile $E_v(x,y)$ visualizes the shape of the V-groove QWR (white) along a $(1\bar{1}0)$ cross section plane, as well as the GRINSCH structure of the barrier (yellow). The pixel size is $1\,\mathrm{nm}^2$, being the spatial resolution of the finite differences grid. (b) Piezoelectric potential along the same cross section plane. The spatial profiles of the strain tensor components are shown along the same plane: (c) ϵ_{xx}, (d) ϵ_{zz}, (e) ϵ_{xy}, and (f) ϵ_{xz}. The resulting effective lateral confinement potentials for electrons and holes are shown in (g) and (h), respectively.

	Unit	Value in $In_cGa_{1-c}As$	Ref.
E_0	meV	$1424 - 1501c + 433c^2$	[69]
E'_v	meV	$-6920 + 231c - 58c^2$	[58,69]
Δ_0	meV	$340 - 93c + 133c^2$	[45]
m_e	m_0	$0.0667 - 0.0419c$	[62]
m_{so}	m_0	$0.163 - 0.023c$	[62,68]
γ_1		$1/[(1-c)/7.10 + c/19.7]$	[45,61]
γ_2		$1/[(1-c)/2.02 + c/8.4]$	[45,61]
γ_3		$1/[(1-c)/2.91 + c/9.3]$	[45,61]
B	meV nm^2	0	[33,240]
b'	meV	0	[33,240]
a_c	meV	$-8013 + 2933c$	[62]
a_g	meV	$-8233 + 2153c$	[62]
b_v	meV	$-1824 + 24c$	[45]
d_v	meV	$-5062 + 1462c$	[45]
C_{11}	GPa	$118.8 - 35.5c$	[45,61]
C_{12}	GPa	$53.8 - 8.5c$	[45,61]
C_{44}	GPa	$59.4 - 19.8c$	[45,61]
ϵ_s		$13.18 + 1.42c$	[62]
e_{14}	C m^{-2}	$0.160 - 0.115c$	[62]

Table 7.1:

Indium gallium arsenide material parameters for 300 K used in this chapter.

$\psi_{y\Uparrow}$, $\psi_{z\Uparrow}$, $\psi_{s\Downarrow}$, $\psi_{x\Downarrow}$, $\psi_{y\Downarrow}$, and $\psi_{z\Downarrow}$ is obtained, analogously to Eq. (4.14). However, Eq. (7.2) and its eigensolutions (E_n, Ψ_n) depend continuously on k_z in the case of a QWR, constituing the 1D bandstructure with k_z as independent variable.

Since the strain in V-groove QWRs is significantly weaker than obtained by Stranski-Krastanow QD growth (in order to enable a dislocation-free growth of the wire) and since the quantization energies in QWRs are generally smaller than in InAs/GaAs QDs (due to the larger size of the QWR) the arguments in Chapter 3 supporting the accuracy of the eight-band $\mathbf{k}\cdot\mathbf{p}$ model doubtlessly hold here, too. Further, the common literature values for the $\mathbf{k}\cdot\mathbf{p}$ parameters provide a good description as the populated range of the 1D-bandstructure does not depart too far from the Γ point. Tables 7.1 and 7.2 display the $\mathbf{k}\cdot\mathbf{p}$ material parameters used here for InGaAs and AlGaAs, respectively, at room temperature.

Fig. 7.2 shows the lowest four 1D CBs and the topmost twelve 1D VBs along the z direction for k_z in the range $\pm\,0.2\,\mathrm{nm}^{-1}$. All subbands were calculated separately, starting with a Gaussian random vector for the wave function Ψ at

	Unit	Value in $\mathrm{Al}_c\mathrm{Ga}_{1-c}\mathrm{As}$	Ref.
E_0	meV	$1428 + 1572c - 553c^2$	[68]
E'_v	meV	$-6920 - 433c - 24c^2$	[66, 68]
Δ_0	meV	$340 - 131c + 71c^2$	[68]
m_e	m_0	$0.067 + 0.083c$	[68]
m_{so}	m_0	$0.165 + 0.135c$	[68]
γ_1		$1/[(1-c)/7.10 + c/3.76]$	[68]
γ_2		$1/[(1-c)/2.02 + c/0.90]$	[68]
γ_3		$1/[(1-c)/2.91 + c/1.42]$	[68]
B	$\mathrm{meV\,nm^2}$	0	[33, 240]
b'	meV	0	[33, 240]
a_c	meV	$-8013 + 2373c$	[58]
a_g	meV	$-8233 + 123c$	[58]
b_v	meV	$-1824 + 124c$	[58]
d_v	meV	$-5062 + 512c$	[58]
C_{11}	GPa	$118.8 + 1.4c$	[68]
C_{12}	GPa	$53.8 + 3.2c$	[68]
C_{44}	GPa	$59.4 - 0.5c$	[68]
ϵ_s		13.46	
e_{14}	$\mathrm{C\,m^{-2}}$	0.137	

Table 7.2:

Aluminium gallium arsenide material parameters for 300 K used in this chapter. The expression for $E_0(c)$ is a second-order Chebyshev approximation to $1424 + 1594c + c(c-1)(1310c - 127)$ for $c \leq 0.45$.

$k_z = 0$ and using the obtained wave function at each k_z value as initial guess for the next k_z value, adjusting the search energy[2] by second order extrapolation. The energy zero in Fig. 7.2 is the CB edge of unstrained bulk $\mathrm{In_{0.2}Ga_{0.8}As}$. The VB states are labeled from top to bottom by $|1,1\rangle$, $|1,2\rangle$, $|2,1\rangle$, $|2,2\rangle$, a. s. o., taking into account the spin projection. The $|\cdot,1\rangle$ states are shown as solid lines and the $|\cdot,2\rangle$ states as dashed lines.

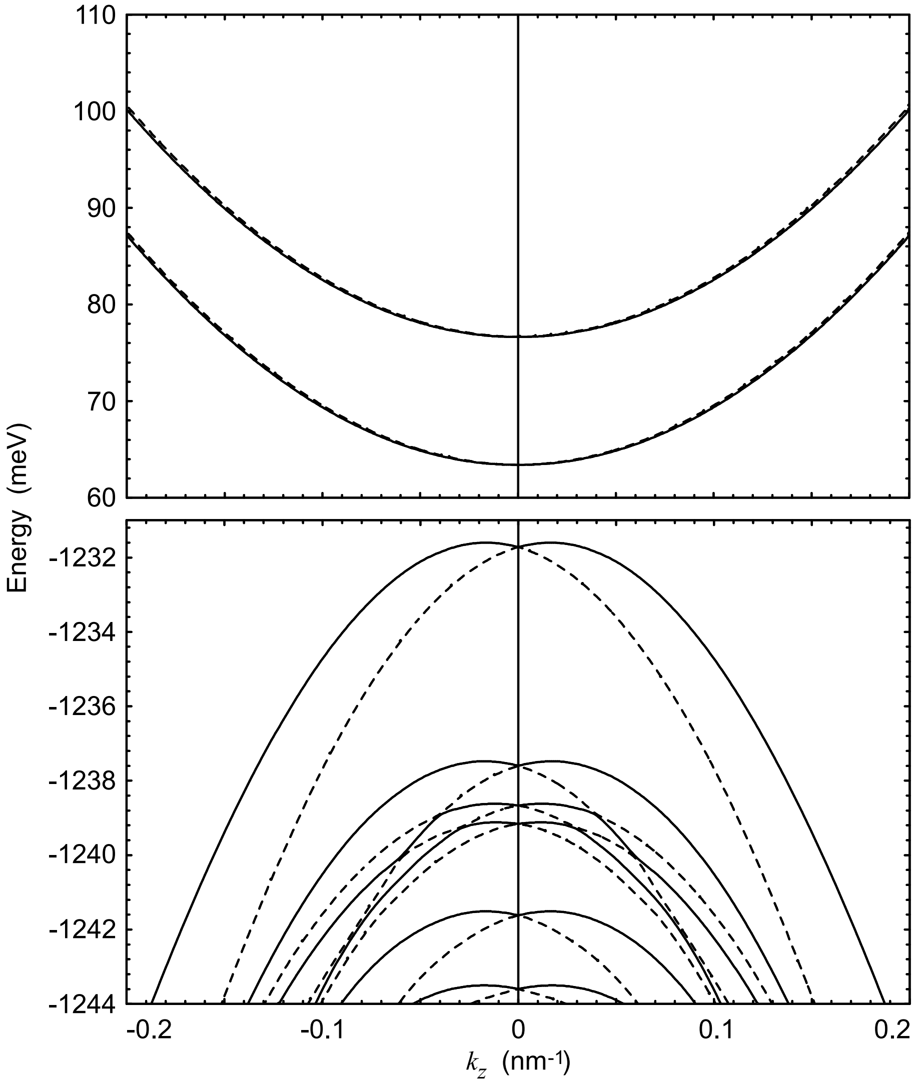


Figure 7.2:

Lowest four 1D CBs and topmost twelve 1D VBs in the QWR (each shown by alternating solid and dashed lines). The energy zero is the CB edge of unstrained bulk $\mathrm{In_{0.2}Ga_{0.8}As}$.

---

[2] See Section 9.3.

Between the $|\cdot,1\rangle$ and $|\cdot,2\rangle$ states a meV-range splitting is observed for $k_z \neq 0$. This spin splitting is caused by the mesoscopic violation of the spatial inversion symmetry due to the irregular geometrical shape of the QWR, the strain field, and the piezoelectric charging. The spin splitting leads to double peaks in the DOS. Each peak doublet consists of a singularity accompanied by a finite discontinuity. This is well seen in the VB-DOS, see Fig. 7.3.

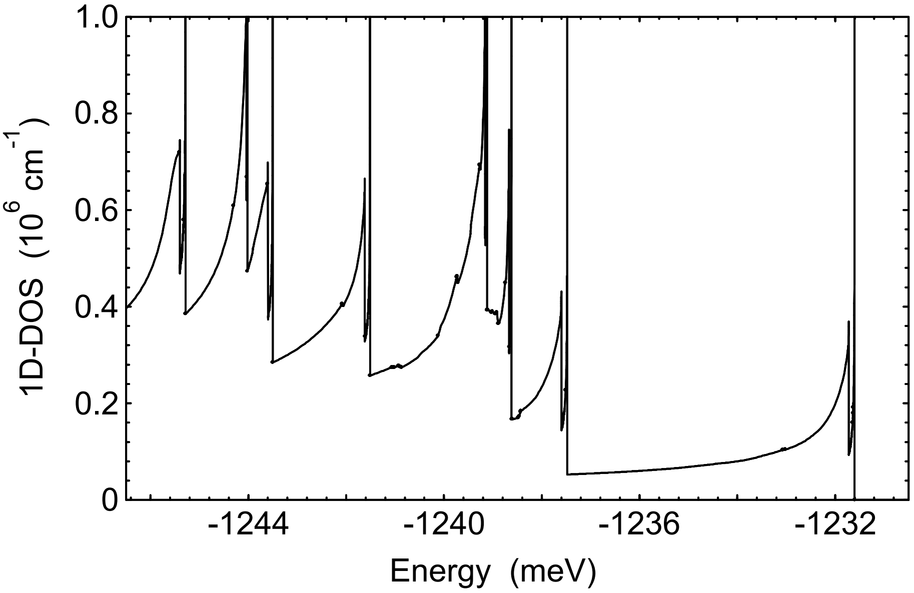


Figure 7.3:

The peak doublets in the VB DOS reflect the spin splitting in the 1D bandstructure.

In general, the spin splitting causes slightly increased homogeneous linewidths of all transitions, as compared to the (hypothetical) unsplit case, and presents a lower limit of the linewidth for disappearing interface roughness. Fig. 7.4 shows the joint DOS for **k**-conserving transitions and visualizes the dominance of the singularities at all subband edges. A weakening of the singularity peaks due to electron-hole Coulomb correlation has been predicted on the basis of generalized semiconductor Bloch equations [241]. The SMEs between $\langle a|$ and spin-split states $|b,1\rangle$, $|b,2\rangle$ may differ by a factor 100 only, reflecting a relaxation of spin-selection rules due to band mixing.

The piezoelectric potential V_p effectively reduces the actually tapered QWR thickness from $< 23\,\mathrm{nm}$ to uniformly $\sim 7\,\mathrm{nm}$. Hence, for holes the QWR looks approximately like a bent quantum well, as can be seen in Fig. 7.1(h).

Eventually, the situation corresponds to the formation of three electronically coupled sub-QWRs for holes along the lower interface, which have separate

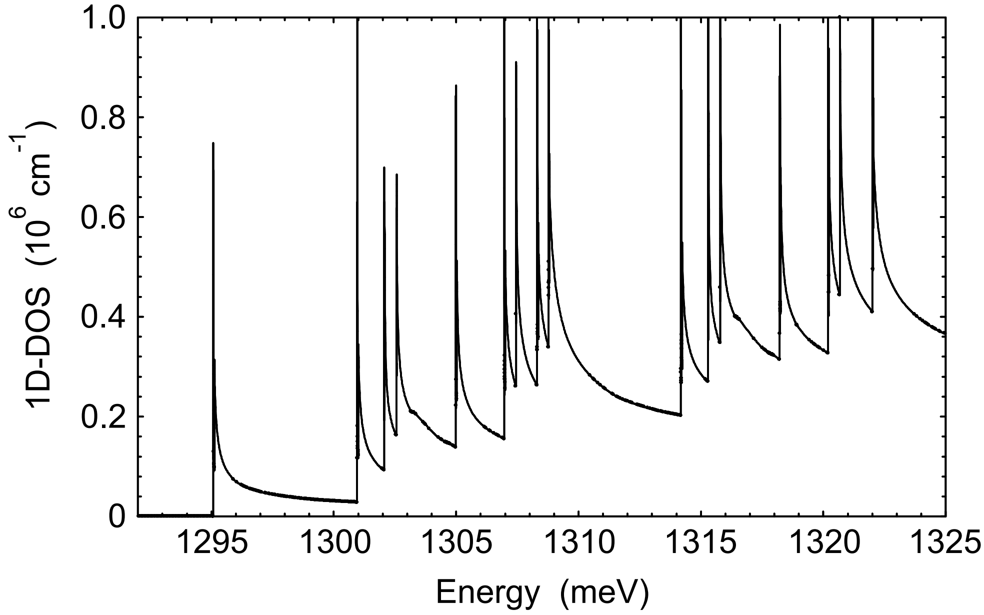


Figure 7.4:

Joint DOS for **k**-conserving transitions (without inhomogeneous broadening).

ground-states at $k_z = 0$, see Fig. 7.5. The lowest of these, the true ground-state $|1,\cdot\rangle$, is located in the center and followed by its first excited state $|2,\cdot\rangle$. The next two states, $|3,\cdot\rangle$ and $|4,\cdot\rangle$, have a $\sim 7\,\mathrm{meV}$ larger quantization energy, while their separation is only $\sim 0.5\,\mathrm{meV}$. This gives rise to a weak left-right symmetry breaking producing the two other quasi-ground-states. Fig. 7.5 shows only the left of them.

A previous study of the symmetry breaking in this QWR neglecting band mixing [239] suggested that only the true ground-state $|1,\cdot\rangle$ is affected, rather than also the second excited state. The eight-band **k·p** calculation discriminates this inaccuracy and indicates that the single-band effective-mass approximation is unsuited also to QWRs in such complex situations involving strain and piezoelectricity within an irregular geometry. The main source of inaccuracy in the present case is that different states, e. g. $|1,\cdot\rangle$ and $|3,\cdot\rangle$, experience different directions of strongest confinement. Simple effective mass models are not flexible enough to resemble this situation in the Hamiltonian. Even the 1D-CB is non-parabolic: For $k_z = 0 \ldots 1.4\,\mathrm{nm}^{-1}$ the ground-state effective mass increases from 0.066 to 0.208, as shown in Fig. 7.6. Due to the large size of the QWR the electron effective mass approximately takes on its bulk value at $k_z = 0$. The increase occurs at higher wave numbers k_z , not until the wave function is more affected by the structural and piezoelectric confinement.

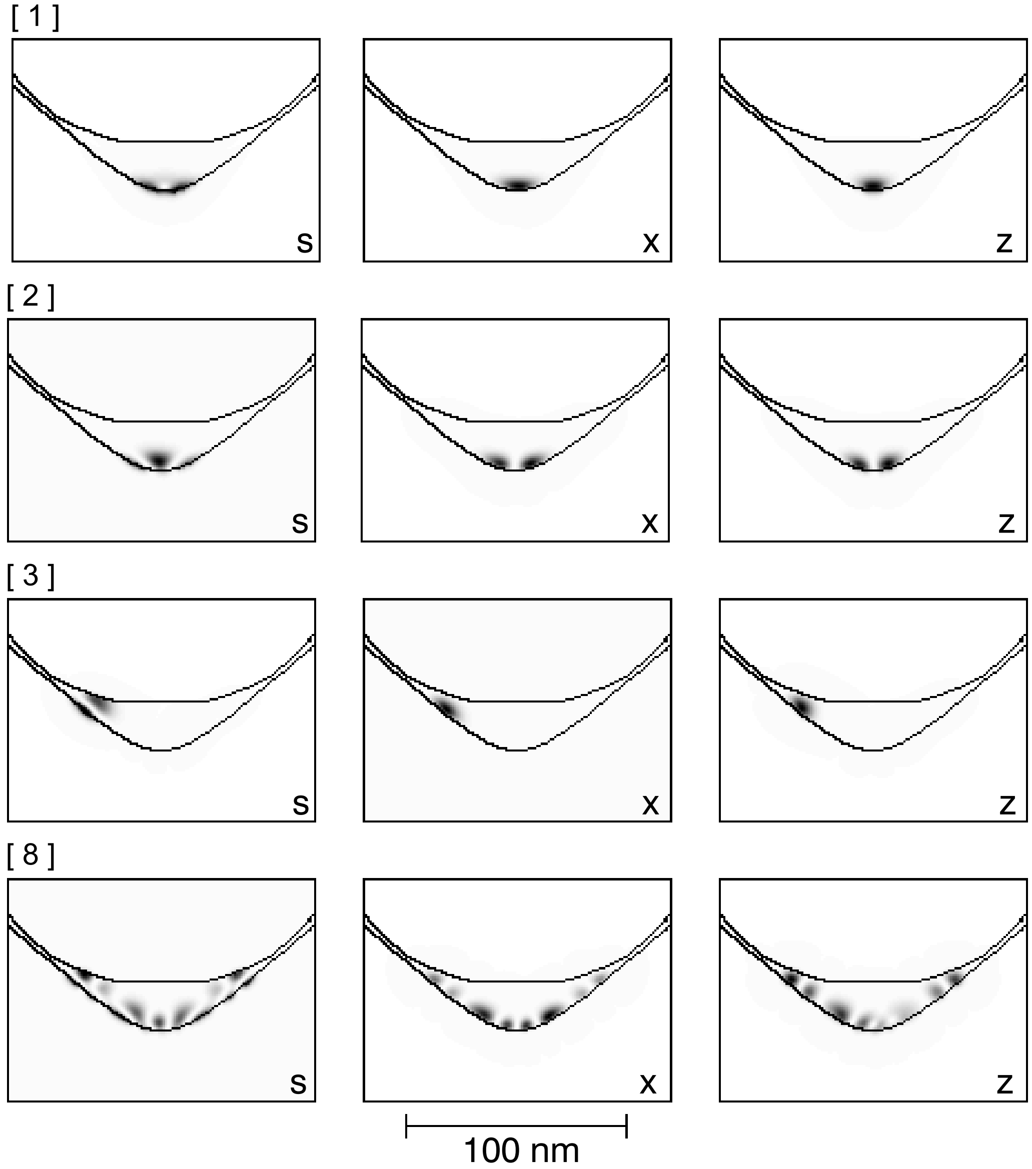


Figure 7.5:

Envelopes $|\psi|^2$ of the Bloch functions $|s\uparrow\rangle$, $|x\uparrow\rangle$, and $|z\uparrow\rangle$ in the VB states $|1,\cdot\rangle$, $|2,\cdot\rangle$, $|3,\cdot\rangle$, and $|8,\cdot\rangle$ at $k_z = 0$, plotted along the $(1\bar{1}0)$ cross section plane.

7.2 Specific quantum wire structures

Analogous to the case of pyramidal QDs, the interplay of structural confinement, strain, piezoelectricity, and band mixing gives rise to very particular optical properties of QWRs, some of which were studied theoretically in the frame of this work.

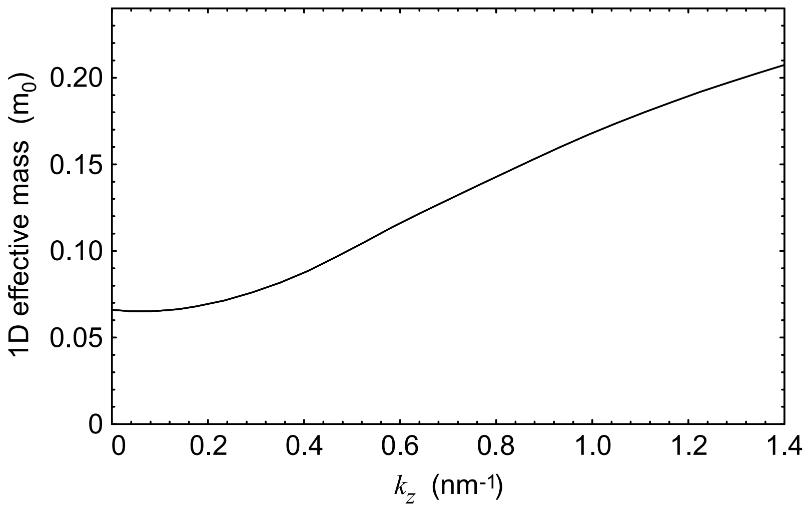


Figure 7.6: Longitudinal effective mass in the CB state $|1,1\rangle$.

7.2.1 $\mathbf{In_{0.15}Ga_{0.85}As/Al_{0.15}Ga_{0.85}As}$ V-groove quantum wires

Fig. 7.7(a) shows polarization dependent PLE spectra of pseudomorphic $\mathrm{In_{0.15}Ga_{0.85}As/Al_{0.15}Ga_{0.85}As}$ V-groove QWRs oriented along $[1\bar{1}0]$ by growth on a prepatterned (001) substrate [16, 228]. Each peak corresponds to the transition from an electron to a hole subband in the QWR. The measured transitions are compared to eight-band $\mathbf{k}\cdot\mathbf{p}$ calculations based on the QWR cross section shape known from transmission electron microscopy and assuming the nominal chemical composition of the wire. The comparison indicates a high accuracy of such calculations. Transitions between different subbands show different polarization anisotropies obeying no plain rules. The influence of the strain and related piezoelectricity on the polarization of the optical absorption in irregularly shaped QWRs can be seen from Fig. 7.7(b) showing the calculated optical anisotropy spectra of a strained and an unstrained QWR having the same cross section shape. A detailed discussion of the optical properties based on the calculations and PLE measurements is given in Ref. [16, 228].

7.2.2 $\mathbf{In_{0.7}Ga_{0.3}As/InP}$ V-groove quantum wires

An extreme case of piezoelectric impact on the optical properties of QWRs was found in $\mathrm{In_{0.7}Ga_{0.3}As/InP}$ V-groove QWRs [226], where a quenching of the ground-state transition luminescence is predicted [225]: The dominant piezoelectric field (in average $3.9\,\mathrm{kVcm^{-1}}$ in lateral direction) effectively partitions the QWR into spatially distinct wires for electrons and holes, forcing a geometrical

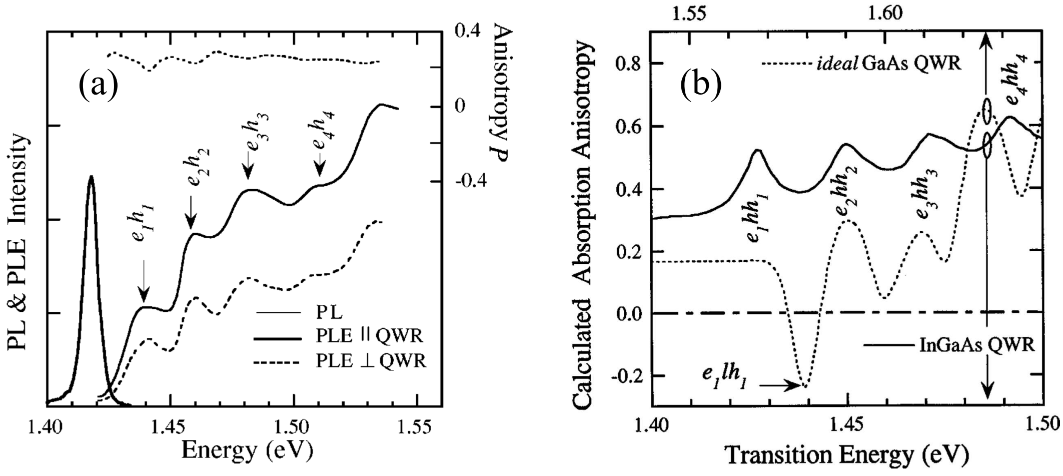


Figure 7.7:

(a) PL, PLE, and polarization anisotropy (P) spectra of pseudomorphic $\mathrm{In_{0.15}Ga_{0.85}As/Al_{0.15}Ga_{0.85}As}$ V-groove QWRs. P is the anisotropy between the polarization directions parallel (solid line) and perpendicular (dashed line) to the QWR direction $[1\bar{1}0]$ in the substrate plane (001). The arrows indicate the 1D subband edge transitions according to eight-band $\mathbf{k}\cdot\mathbf{p}$ calculations based on the actual QWR cross section geometry known from transmission electron microscopy. (b) Calculated anisotropies P of the strained QWR and of an ideally lattice-matched $\mathrm{GaAs/Al_{0.15}Ga_{0.85}As}$ QWR having the same cross section shape. The figures are taken from Ref. [16].

symmetry breaking of the uppermost 1D valence subband and realizing spatially indirect transitions similar to the situation in type-II QWRs [128, 129]. At moderate optical excitation ($< 100\,\mathrm{Wcm^{-2}}$) the injected carriers ($< 5 \times 10^5\,\mathrm{cm^{-1}}$) are estimated to screen the field to $\geq 2.6\,\mathrm{kVcm^{-1}}$. This is only a weak screening of the piezopotential so that the spatial electron-hole separation is not suspended, enabling an experimental observation of the reduced ground-state luminescence intensity. The study was published in Ref. [225].

7.2.3 Strained cleaved-edge overgrowth quantum wires

By cleaved-edge overgrowth, type-I QWRs can be fabricated which confine charge carriers without providing a closed 2D heterojunction barrier. The confinement in such QWRs can be enhanced by inhomogeneous strain, e. g. in the material system $\mathrm{In_{0.20}Al_{0.80}As/Al_{0.35}Ga_{0.65}As/GaAs}$. In contrast to less advanced bandstructure models, eight-band $\mathbf{k}\cdot\mathbf{p}$ theory predicts localization energies of up to 90 meV in strained cleaved-edge overgrowth QWRs, as detailed in Ref. [242].

Part III

Numerical Procedures

Chapter 8

Finite Differences Method

To generate matrix representations H of the Hamiltonians Eq. (4.14) and Eq. (7.2) a finite differences method (FDM) on a cubic grid is used [40], having $\tilde{N} = N_x \cdot N_y$ nodes in the case of a QWR and $\tilde{N} = N_x \cdot N_y \cdot N_z$ in the case of a QD. The real and imaginary parts of the envelope functions are represented by $\tilde{N}$-dimensional vectors, each. Accordingly, the differential operators and spatially varying parameters in Eq. (2.1) become $\tilde{N} \times \tilde{N}$ real matrices, so that H is an $8\tilde{N} \times 8\tilde{N}$ complex matrix and the (real) dimension of a complete wave-function vector is $N = 16\tilde{N}$. The $\tilde{N} \times \tilde{N}$ submatrices representing material parameters or strain components are diagonal. The matrix representations of the differential operators replacing the **k** components in Eq. (2.1) are chosen as to reproduce the symmetry properties and commutation relations of the **k** components, while assuring the numerical stability of the difference scheme.

This is accomplished by introducing two different matrices, $D^{(+)}$ and $D^{(-)}$, for each partial derivative (∂_x , ∂_y , and – if applicable – ∂_z), whereby

$$[D^{(+)}]^T = -\,D^{(-)} \qquad .$$

The $D^{(+)}$-type differentiator matrices are used in the upper right triangular matrix of H and the $D^{(-)}$-type differentiator matrices in the lower left triangular matrix. The discretizations of the symmetrized expressions for $Q\partial_x$, $Q\partial_y$, $Q\partial_z$, $Q\partial_{xx}$, $Q\partial_{xy}$, $Q\partial_{xz}$, $Q\partial_{yy}$, $Q\partial_{yz}$, $Q\partial_{zz}$ (see Section 2.2) are obtained from the standard first order forward (backward) difference quotients ${\partial_x}^{(+)}$, ${\partial_y}^{(+)}$, ${\partial_z}^{(+)}$ (${\partial_x}^{(-)}$, ${\partial_y}^{(-)}$, ${\partial_z}^{(-)}$) by

$$\begin{aligned}
D_x{}^{(+)}(Q) &= \left[Q\partial_x{}^{(+)} + \partial_x{}^{(+)}Q\right]/2 \quad ,\\
D_x{}^{(-)}(Q) &= \left[Q\partial_x{}^{(-)} + \partial_x{}^{(-)}Q\right]/2 \quad ,\\
D_y{}^{(+)}(Q) &= \left[Q\partial_y{}^{(+)} + \partial_y{}^{(+)}Q\right]/2 \quad ,\\
D_y{}^{(-)}(Q) &= \left[Q\partial_y{}^{(-)} + \partial_y{}^{(-)}Q\right]/2 \quad ,\\
D_z{}^{(+)}(Q) &= \left[Q\partial_z{}^{(+)} + \partial_z{}^{(+)}Q\right]/2 \quad ,\\
D_z{}^{(-)}(Q) &= \left[Q\partial_z{}^{(-)} + \partial_z{}^{(-)}Q\right]/2 \quad ,\\
D_{xx}(Q) &= \left[\partial_x{}^{(+)}Q\partial_x{}^{(-)} + \partial_x{}^{(-)}Q\partial_x{}^{(+)}\right]/2 \quad ,\\
D_{yy}(Q) &= \left[\partial_y{}^{(+)}Q\partial_y{}^{(-)} + \partial_y{}^{(-)}Q\partial_y{}^{(+)}\right]/2 \quad ,\\
D_{zz}(Q) &= \left[\partial_z{}^{(+)}Q\partial_z{}^{(-)} + \partial_z{}^{(-)}Q\partial_z{}^{(+)}\right]/2 \quad ,\\
D_{xy}(Q) &= \left[\partial_x{}^{(+)}Q\partial_y{}^{(+)} + \partial_y{}^{(+)}Q\partial_x{}^{(+)} + \partial_x{}^{(-)}Q\partial_y{}^{(-)} + \partial_y{}^{(-)}Q\partial_x{}^{(-)}\right]/4 \quad ,\\
D_{xz}(Q) &= \left[\partial_x{}^{(+)}Q\partial_z{}^{(+)} + \partial_z{}^{(+)}Q\partial_x{}^{(+)} + \partial_x{}^{(-)}Q\partial_z{}^{(-)} + \partial_z{}^{(-)}Q\partial_x{}^{(-)}\right]/4 \quad ,\\
D_{yz}(Q) &= \left[\partial_y{}^{(+)}Q\partial_z{}^{(+)} + \partial_z{}^{(+)}Q\partial_y{}^{(+)} + \partial_y{}^{(-)}Q\partial_z{}^{(-)} + \partial_z{}^{(-)}Q\partial_y{}^{(-)}\right]/4 \quad .
\end{aligned}$$

Expanding these expressions yields the actual difference quotients used in this work. For Q being the identity matrix all of them reduce to standard difference quotients. The handling of boundary nodes for the realization of Neumann conditions is standard.

Chapter 9

Calculation of Confined States

Quantum confined structures, like QWRs and QDs, differ from bulk material by their finite spatial structure destroying the translation symmetry of the crystal lattice in, at least, one spatial direction (two for QWRs, three for QDs). Bandstructure models can take this into account either on a microscopic scale, like the tight-binding and pseudopotential methods, or on a mesoscopic scale, like the envelope function approximation (EFA) methods. Regardless of the particular nanostructure or theoretical concept, modeling approaches for the optical properties of semiconductor heterostructures share the necessity to describe both CB and VB states, having to construct a Hamiltonian containing both the CB and VB dispersions (or level structure) of the low-dimensional structure.

The requirement to expand the Hamiltonian into a finite basis for practical calculations (discretization) introduces energy bounds ("cut-off") to the Hamiltonian and leads to the existence of a "maximally excited CB state" at extremely high energy and a "maximally excited VB state" at extremely low energy. These states are not always physically realistic, but the associated upper and lower bounds of the discrete spectrum lay far beyond the quantization energies of the wanted, confined states. The desired wave functions (eigenvectors) must therefore be calculated without knowledge about the extreme eigenvalues and -vectors of the Hamiltonian. This situation differs from, e. g., the solution of the Kohn-Sham equations within density functional theory (DFT) where the groundstate properties of a system are constructed from eigenvectors corresponding to the lowest eigenvalues of a Hamiltonian employing the self-consistent, equilibrium potential.

Within envelope function theory there are common approaches to decouple the CB and VB parts of the Hamiltonian, either by treating their interaction perturbatively [23] or by neglecting it. Representants of the latter category are the four-band model [24, 174, 243, 244] and the six-band model [172, 173] for VB states, and the (single band) effective mass approximation [48, 54, 55] originally derived for CB states, but also applied to VB states. These calculation schemes offer the computational advantage that the confined CB (VB) states of interest appear at the lower (upper) spectral edge of the Hamiltonian, being associated to *outermost* eigenvalues. This facilitates their calculation substantially.

Partial justification for such approaches results from the fact that neglecting s-type Bloch functions in calculations of confined VB states produces inaccuracies often being tolerable (see Chapter 3). On the other hand, the neglection of p-type Bloch functions in calculations of confined CB states causes more significant errors. CB groundstate calculations based on decoupled bandstructure models for the bulk materials can be improved by a perturbative inclusion of the remote band effects originating from the VB. However, the accuracy of excited states calculated by such methods is limited [23]. Hence, the CB-VB interaction should be included not only for nanostructures formed from narrow-gap semiconductors, like InAs, but also for wide-gap semiconductor materials, like InP, GaAs, or CdSe.

This chapter introduces the algorithm developed in the frame of this work to solve the algebraic eigenproblem arising from the eight-band **k·p** model applied to low-dimensional structures. It is emphasized that ready-to-run software for such problems is extremely rare still today, let alone, five years ago, and that even the theoretical background for the development of such procedures advances well beyond textbook knowledge of numerical mathematics.

The algorithm proposed here is designed for the efficient calculation of interior eigenvalues and -vectors of Hamiltonians with dimensions $N \sim 10^7$ without reference to other eigenpairs. It is related to prevailing methods for the same purpose [245–247] but performs faster in all present applications. The algorithm is valuable for theoretical investigations of nanostructures also beyond this work, as it is applicable to other kinds of Hamiltonians as well, for instance those arising from tight-binding or pseudopotential methods.

The principal question having stimulated the achievements reported in this chapter was: *Which is the fastest method for calculating confined states from the given eight-band* **k·p** *Hamiltonian for QWRs or QDs?* The answer could only be found by (i) a scrutiny of the available literature (taking into account advanced, recently developed methods used in fields like DFT, EP theory, and quantum chemistry) and (ii) seeking for options to improve the most promising candidate.

Such improvement was necessary since most of the known procedures designed for the calculation of interior eigenvalues are inapplicable to the discrete eight-band **k·p** Hamiltonian because of its exceedingly large rank. On the other hand, the "folded spectrum" method successfully employed in (mathematically) comparable EP calculations displays a potentially important draw-back, presenting a challenge for improvement. Ideas for such improvement could be found by analyzing a variety of iterative eigensolution methods, taking also into account procedures for the calculation of extreme eigenvalues.

The main obstacle encountered in the comparison of procedures introduced or employed in more than 50 works is that benchmarks are rarely reported, and that their results, if existing, are transferable to other Hamiltonians or calculation platforms only with caution. Hence, the effiency of reported procedures had to be estimated based on their mathematical properties only, requiring a coarse classification scheme for basic algorithm types, and deliberate individual analysis for numerous modifications within each category. Plenty of interrelations between procedures, as identified by means of numerical analysis, eventually help to organize the wealth of methods within a hierarchic system, which is outlined in Section 9.2.

By the nature of the matter, the numerical analysis of recent, advanced procedures derived from different algorithmic base types, is of pronouncedly technical kind. Moreover, the presentation is condensed, taking care of its secondary role within the frame of this work. Although the analysis in Section 9.2 was essential for the accomplishment of the present work, it is not prerequisite for an *adaption of the proposed algorithm to other work*. Therefore, readers seeking information only on the structure and implementation of the improved algorithm, or its efficiency, are encouraged to skip Sections 9.1 and 9.2, and to continue on p. 156.

The following section recapitulates fill-in structures of Hamiltonian representations arising in different bandstructure calculation methods, in order to motivate the iterative ansatz underlying the algorithms relevant to this work.

9.1 Hamiltonian fill-in structures

In the semi-empirical tight-binding model [35, 183, 248–253] the Hamiltonian $\hat{H}$ for the heterostructure is expanded into orbital functions ζ_i $(i = 1, \ldots, N)$ each of which is assigned to a certain atom. Since only interactions between near(est) neighbor atoms are considered each basis function ζ_i yields non-zero matrix elements $h_{ij} =< \zeta_i | \hat{H} \zeta_j >$ only with a small, and essentially constant, number of other functions ζ_j. Thus the overall number of non-zero matrix elements of $\hat{H}$ is proportional to N instead of N^2, and $H = (h_{ij})$ possesses a sparse matrix representation. If large numbers of atoms are considered, e. g., 4047 for one free-standing InAs QD [35], the rank is $N = 20235$, or even larger in other examples.

The EFA can be carried out in any representation of $\hat{H}$, in particular the real ($\mathbf{r}$-)space or the reciprocal ($\mathbf{k}$-)space. The most natural formulation of a heterostructure refers to the $\mathbf{r}$-space and leads to a set of coupled partial differential equations for the envelope functions Ψ, like Eq. (4.14) or Eq. (7.2). From such sets a few eigensolutions shall be extracted, specified by an interval of interest for the eigenvalues. The goal is, in any representation, achieved by expanding $\hat{H}$ into a basis $\mathcal{B}$ of N functions $\zeta_i(\mathbf{r}[,\mathbf{k}])$ (discretization) yielding an algebraic, generalized eigenproblem $H\Psi = \lambda G\Psi$ where $G = \mathcal{B}^\dagger\mathcal{B}$ is the metric tensor of the basis and $H = \mathcal{B}^\dagger\hat{H}\mathcal{B}$ is the projection of $\hat{H}$ onto span$(\mathcal{B})$.

By the discretization $\hat{H}$ is mapped to $G^{-1}H$. This is no equivalence transformation unless a complete, i. e. infinite, orthonormal basis $\mathcal{B}$ is used. In practice, one always uses finite, i. e. incomplete, and often even non-unitary bases $\mathcal{B}$, so that $\hat{H}$ and its discrete counterpart have different eigenvalues, and the discrete eigenvector approximations are deflected from orthogonal projections of the eigenvectors to the finite subspace. The most common non-unitary basis sets are those arising in finite differences methods (FDM). As the basis size N approaches infinity, $[G,H] \to [0]$ so that the discretization asymptotically approaches an equivalence transformation. Therefore, G is usually neglected also for finite N, i. e., replaced with the identity matrix I. The true eigensolutions of $\hat{H}$ are obtained as the asymptotical solutions $(N \to \infty)$ of the discrete problem, as in the FDM, the finite elements method (FEM), or Fourier type methods (FM). The FDM *additionally* requires that the discrete expressions for differentiations are asymptotic approximants for the respective differential operators.

The feasible choices of $\mathcal{B}$ for multi-band EFA Hamiltonians may be categorized according to the spatial overlap of the basis functions $\zeta_i(\mathbf{r}[,\mathbf{k}])$. Plane waves overlap in the entire **r**-space, while finite elements have common boundaries only, and the basis functions underlying the FDM share single grid points only. The plane wave expansion realizing the **k**-representation and used in FMs yields at most non-zero matrix elements h_{ij} due to **r**-dependent terms, resulting in dense projections H. In contrast, the FEM and FDM both produce sparse matrices H of much larger rank than those obtained by FM.

The expense of handling dense matrices is prohibitively progressive in N: The storage demand and the (floating point) operation count[1] (OC) for matrix vector multiplications scale with N^2, while N is antiproportional to the spatial resolution d represented by $\mathcal{B}$ in *each* spatial direction. Accordingly, for two- or three-dimensionally confined structures the exponent of this power-law dependence is 4 or 6, respectively, soon becoming untractable. Hence, the main application field of FM in the frame of the EFA are quantum well problems (*pars pro toto* see Ref. [33, 255]) although also a few EFA+FM calculations of QWRs were reported [59, 173, 231, 239, 256]. The majority of EFA based studies of QWRs and QDs used the FEM [244, 257–261] or FDM [31, 40, 114, 128, 148, 207, 262–267]. Other studies took advantage of the sparse-matrix eigenproblems resulting from the tight-binding [35,248,250–252] or effective bond-orbital model (EBOM) [268,269] approaches.

---

[1] The traditional neglection of integer operations aside of floating point operations in complexity estimates is no more justified for the contemporary processor architectures (e. g., the Intel Pentium, AMD K, or DEC Alpha families) allowing the execution of both operation types, even simultaneously, within one clock cycle [254].

Pseudopotential methods also benefit from using plane waves, however, without involving dense matrices. Relying on the plane-wave expansion, matrix vector multiplications $\hat{H}\Psi$ can be split into a **k**-space sparse multiplication and a **r**-space sparse multiplication, connecting both by fast Fourier transformation [270]. Using this spectral method (SM) H is split and its parts are stored and applied to Ψ within the different representations [267, 271]. Pseudopotential based DFT calculations addressing the minimization of the total energy functional can be separated into an eigenproblem and a simultaneous self-consistency problem. Common solution methods [270, 272] therefore nest different iteration procedures for improving the eigenvector approximation and for calculating the self-consistent Kohn-Sham potential. Hence, total energy minimization problems are an additional field requiring powerful eigensolution algorithms [273].

The tractability of a large eigenproblem clearly requires avoiding dense matrices (OC scales with N^2) which not necessarily implies the use of sparse matrices (OC scales with N) since there are representation-mixed evaluation schemes for $H\Psi$ (OC scales with $N \log_2 N$). The only remaining criterion is that multiplications $H\Psi$ can, in whatsoever manner, be calculated sufficiently fast. In the following we will therefore refer to (linear) operators rather than matrices H. As for the rest, self-consistency is not addressed.

The above considerations show that the calculation of bandstructures in semiconductor nanostructures almost naturally leads to the extraction of $\ll N$ eigensolutions, specified by an interval for the eigenvalue, from Hermitian operators H with a rank $N \gg 10,000$. These eigenvalues are not located near the upper or lower bound of the spectrum of $\hat{H}$ and must, hence, be calculated *without knowledge about other eigenstates*, due to the large rank.

9.2 Efficiency of prevailing eigensolvers

For the calculation of outermost (extreme) eigenvalues of very large matrices plenty of procedures have been proposed and it is, of course, beyond the scope of this work to review them (see Ref. [270, 274–278] for an introduction). Descriptions of such algorithms are widely scattered throughout the literature on chemical physics, solid states physics, and computational physics / chemistry while only some of them are considered in reviews or textbooks. This section analyzes mathematically both basic procedures and auxiliary tools, quantifying the computational efficiency on the basis of the asymptotic convergence rate. An overview over the various methods is partly obscured by

(i) auxiliary tools to improve existing procedures (e. g., new preconditioners or specific implementations) and

(ii) modifications prior to the actual eigensolver, like (a) multigrid techniques [180] and (b) eigenvalue economization methods partitioning a problem in advance according to its geometrical, mechanical, dynamical, etc., structure (e. g., finite elements, component mode synthesis, and singular perturbation, respectively) [276].

The effort of this chapter is dedicated to the reduction of run time by accelerated convergence. We concentrate on the actual eigensolver, taking only marginally into account achievements from category (i) as these, generically, are of very specialized character and strongly depend on the given operator H. The whole variety of procedures can be ordered according to a hierarchy enabling their straight-forward definition and comparison. As universal performance measure we consider the asymptotic convergence rate of a trial vector towards an eigenvector at the (final) dimension N of the discrete problem: Though approximations for the (finally) desired discrete eigenpair can be obtained by preliminary calculations at a lower dimension (e. g., within multigrid schemes) the final diagonalization within the full rank subspace still causes a significant, if not the largest, contribution to the overall expense. Any concept from (i) and or (ii) may provide *additional improvement* to the basic methods considered in the following, but will hardly contravert the performance ranking dominated by the capabilities of the basic eigensolvers.

Regarding the structure and large rank N of the discretized eight-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian we generally assume below that any suitable algorithm refers to H exclusively by forming matrix vector products $H\mathbf{x}$ while being *unable to transpose, factorize, or invert* H or significantly large submatrices thereof. Furthermore, we assume that the expense to calculate $H\mathbf{x}$ dominates over all other operations so that the most efficient algorithm needs the smallest number of such multiplications. To construct an eigenvector $\mathbf{u}$ iterative algorithms exploit the stationarity of the Rayleigh quotient $\rho = (\mathbf{u}^\dagger H\mathbf{u})/(\mathbf{u}^\dagger\mathbf{u})$ or of the residual norm $||H\mathbf{u} - \rho\mathbf{u}||$.

9.2.1 Ordering scheme for iterative eigensolvers

The direct minimization methods [180, 279] obey a known hierarchy with respect to their efficiency:

preconditioned conjugate-gradient (PCG) >

conjugate gradient (CG) >

gradient >

direction set >

coordinate relaxation methods.

So do the other algorithms used in previous bandstructure calculations:

generalized Davidson algorithm (GDA) >

Lanczos algorithm (LA) >

Chebyshev iteration (CI) >

simultaneous iteration >

subspace iteration >

The two enumerations mix basic algorithms with preconditioning techniques while other methods, like preconditioned gradients or the preconditioned LA, are missing. They are considered in Section 9.2.12.

The two above sets of methods meet on a common level defined by the CG and LA which are equivalent in exact arithmetic. Hence we consider only the CG and PCG algorithms being the most frequently used minimization methods in the field. The second set of iterative methods can be subdivided into (i) methods working with *sequences* of trial vectors and (ii) methods working upon *spaces spanned by sequences* of trial vectors. Generally, the latter perform better than the first. The methods from category (i) can be realized as two- or three-term recurrences while the methods from category (ii) need at least three-term recurrences, however, mostly multi-term recurrences with varying depth. This reflects a more efficient accumulation of information, particularly in presence of finite-precision arithmetic.

Although, from a differential geometry point of view, the two method sets differ not much, the second set is better suited for the spectral description favored here. Therefore the second set is taken as example for explaining how, and to which degree, information on an eigenvector is acquired, used, and maintained by the respective algorithms. To establish the terminology used in this chapter we consider in the following the most elementary iterative method for eigenvector calculation, which at the same time lays the fundament for all advanced algorithms to be considered.

9.2.2 Power method

Consider a non-singular, Hermitian operator H of dimension and rank N with the real eigenvalues $\lambda_1, \ldots, \lambda_N$ in ascending order and the corresponding normalized eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$. Usually H is the discretized Hamiltonian. Only for the sake of simplicity we assume that all eigenvalues are non-degenerate. However, the procedures to be discussed cope with degenerate eigenvalues as well, in particular the simultaneous variants, whereby only some actual convergence rates will change in an unspectacular way. $\Lambda = \lambda_N - \lambda_1$ denotes the *spread* of H. Since H is self-adjoint a normalized trial vector $\mathbf{x}$ can be expanded into the eigenvectors of H, $\mathbf{x} = \sum_{j=1}^{N} c_j \mathbf{u}_j$. Then the multiplication $H\mathbf{x}$ yields $\mathbf{y} = \sum_{j=1}^{N} \lambda_j c_j \mathbf{u}_j$. Hereby either c_1 or c_N is amplified most, depending on whether $|\lambda_1| > |\lambda_N|$ or *vice versa*, so that many successive multiplications (with subsequent normalization each) will convert $\mathbf{x}$ either into $\mathbf{u}_1$ or $\mathbf{u}_N$. This is the fundamental mechanism of *gaining* information on an eigenvector, and the particular procedure is known as vector iteration or *power method* (PM) [277]. We consider a trial vector to contain "new information" if it is linearly independent of all previous trial vectors, regardless how it is constructed. By this definition, gain of information is not conditional on an actually done multiplication with H, which becomes relevant in Section 9.2.14.

If by appropriate gauge $|\lambda_2| \geq |\lambda_N|$ then for large step counts s

$$||\delta\mathbf{x}^{(s)}|| \;=\; ||\mathbf{x}^{(s)} - \mathbf{u}_1|| \;\leq\; C \cdot \kappa^s \;=\; C \cdot |\lambda_2/\lambda_1|^s \tag{9.1}$$

[274], i. e. the normalized vectors $\mathbf{x}^{(s)}$ converge linearly to $\mathbf{u}_1$. For linear convergence we define the *effective asymptotic convergence rate* ϕ by $\kappa^{1/(z\phi)} = 1/\mathrm{e}$ where z is the number of $H\mathbf{x}$ multiplications required to construct a new trial vector, i. e., a measure for the complexity of an iteration cycle (in most procedures $z = 1$). Asymptotic means that we consider ϕ as $s \to \infty$. Throughout this chapter we use ϕ as the universal measure of performance. For the PM ($z = 1$)

$$\begin{aligned} \phi \;&=\; -\frac{1}{z}\ln\kappa \;\approx\; (1-\kappa)/z \;=\; 1 - |\lambda_2/\lambda_1| \\ &<\; 2\gamma \qquad \text{where} \qquad \gamma \;=\; (\lambda_2 - \lambda_1)/\Lambda \end{aligned} \tag{9.2}$$

is the *gap ratio* of λ_1. ϕ depends on the energy gauge [277] and is maximal if $(\lambda_2 + \lambda_N)/2$ is chosen to be the energy zero, as will be shown in Section 9.2.7.

ϕ is related to the evolution of the quantity $\delta\mathbf{x}$, i. e. to the convergence of the eigen*vector*, which is most interesting in practice. Nevertheless, convergence factors κ often refer to the Rayleigh quotient instead, i. e. to the convergence of the eigen*value*,

$$\rho^{(s)} = \rho_H[\mathbf{x}^{(s)}] = <\mathbf{x}^{(s)}|H\mathbf{x}^{(s)}>/||\mathbf{x}^{(s)}||^2$$

or to the residual

$$\mathbf{r}^{(s)} = H\mathbf{x}^{(s)} - \rho^{(s)}\mathbf{x}^{(s)} \qquad .$$

For small $||\delta\mathbf{x}^{(s)}||$, i. e. for large s, those values can be translated into the κ values defined by Eq. (9.1):

$$\begin{aligned} \text{if} \quad |\rho^{(s)} - \lambda_1| \quad &\leq \quad C_\rho \cdot {\kappa_\rho}^s \quad \text{then} \quad \kappa = \sqrt{\kappa_\rho} \quad , && (9.3) \\ \text{if} \quad ||\mathbf{r}^{(s)}|| \quad &\leq \quad C_r \cdot {\kappa_r}^s \quad \text{then} \quad \kappa_r \leq \kappa \leq \sqrt{\kappa_r} \quad . && (9.4) \end{aligned}$$

In many practical cases $\kappa \approx \kappa_r$ holds much better in Eq. (9.4) than the upper bound.

The performance of the PM is characterized by the convergence rate ϕ according to Eq. (9.2). Λ depends explicitly on the energetic cut-off inherent to the chosen basis function set $\mathcal{B}$ and is usually much larger than the energy level separation $\lambda_2 - \lambda_1$ between the ground state (λ_1) and the first excited state (λ_2) in the nanostructure , so that $\gamma \ll 1$, and the PM will converge extremely slow.

We note that ϕ does *not depend on the rank* N of H but only on its *spectral structure*.[2] This is common to all iterative algorithms. Consequently, the behavior of an iteration cannot be predicted on the basis of the rank or the fill-in structure of the matrix representation H. The latter change dependent on the discretization type while the spectral structure of H, characterized by the gap ratio γ, may remain unaltered or change in a different way.

9.2.3 Spectral transformation

Eq. (9.2) shows that the properly shifted (gauged) PM can be accelerated by increasing γ. The gap ratio γ of the multiplication operator in the PM will change if H is replaced by some operator function $A = f(H)$ since f directly affects all eigenvalues. f must be chosen such that $\mathbf{u}_1$ is (at least approximately) an eigenvector also of A and the corresponding eigenvalue μ_1 of A remains extreme. We define the eigenvalue mapping due to f by $\mu_j = f(\lambda_j)$ so that the μ_j in general are not in ascending order. If the evaluation of $A\mathbf{x}$ requires an expense equivalent to z $H\mathbf{x}$ multiplications then γ will have to increase by more than a factor z in order to enhance the effective convergence rate of the PM.

Here we consider only analytic functions f, so that A is Hermitian and commutes with H (non-analytic functions f will be considered from Section 9.2.10

---

[2] However, when the FDM is used for the eight-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian both the spread Λ and the rank N of H depend on the grid spacing d, leading to the implicit relations $\Lambda \propto N$ and $\phi \propto N^{-1}$ for QWR calculations, and $\Lambda \propto Nd$ and $\phi \propto N^{-1}d^{-1}$ for QD calculations.

on). Then only integer rational functions f are available since (by our assumption) the inversion or complete factorization of H are impossible. Given the polynomial degree z, the best polynomial f maximizes the ratio

$$\frac{\min_{j>1} |\mu_1 - \mu_j|}{\max_{i,j>1} |\mu_i - \mu_j|} = \frac{\gamma_A}{1-\gamma_A}$$

which is equivalent to maximizing the gap ratio γ_A of μ_1. In order to possess the desired min-max-property [274] f must be chosen to be the suitably shifted Chebyshev polynomial T_z , and then

$$\begin{aligned} \gamma_A &= z^2\gamma - \frac{2}{3}z^2(z^2-1)\gamma^2 \quad + \mathcal{O}(\gamma^3) \qquad \text{and} \\ \phi &\approx \frac{2}{z}\gamma_A \quad \approx \quad 2\,z\,\gamma \qquad . \end{aligned} \tag{9.5}$$

Thus the acceleration of the convergence of the PM is proportional to z. The method to replace H with $T_z(\alpha H + \beta I)$ with fixed z and appropriate α, β in the PM is known as *Chebyshev accelerated power method* [274, 277]. Obviously, the larger z, the less iteration steps of the PM will be necessary. Thus, in absence of an upper bound for z the ultimate Chebyshev acceleration is to make z so large that only one step of the PM remains necessary. This idea leads to a qualitatively different method which we refer to as Chebyshev iteration, and which will be discussed in Section 9.2.8.

Another function f appreciated by practitioners is the exponential [148, 150],

$$A = \exp(-\alpha H) + \beta I \qquad (\alpha > 0) \tag{9.6}$$

adopted from algorithms to solve the time-dependent Schrödinger equation [267, 280]. The fastest convergence of an exponentially transformed PM is achieved at the shift

$$\beta = -\frac{1}{2}\left[\mathrm{e}^{-\alpha\lambda_2} + \mathrm{e}^{-\alpha\lambda_N}\right] \tag{9.7}$$

making the iteration twice as fast as $\beta = 0$. The idea behind the use of Eq. (9.6) is that the exponential can be evaluated *approximately* by splitting H according to $H = H_1 + \ldots + H_k$ and using $\exp(-\alpha H) \approx \exp(-\alpha H_1) \ldots \exp(-\alpha H_k)$. The error of this approximation due to the possible non-commutativity of the H_i may be as small as $\mathcal{O}(\alpha^3)$. The non-diagonal among the factors $\exp(-\alpha H_i)$ can be evaluated by Cayley's form [180],

$$\exp(-\alpha H_i) \approx \left[I + \frac{\alpha}{2}H_i\right]^{-1}\left[I - \frac{\alpha}{2}H_i\right] \quad + \mathcal{O}(\alpha^2) \qquad ,$$

if H_i can be inverted at moderate expense [148, 267]. In the optimal case the overall expense of evaluating Eq. (9.6) is equivalent to $z \approx 1$. Since Eq. (9.6) produces the gap ratio

$$\gamma_A = \gamma \left[1 + \frac{\alpha}{2}(\lambda_N - \lambda_2) + \frac{\alpha^2}{12}(\lambda_N - \lambda_2)(\lambda_N - 2\lambda_2 + \lambda_1) \quad + \mathcal{O}(\alpha^3) \right]$$

$\alpha \approx 2/\Lambda$ will yield an acceleration of the PM by a factor 2 (for $z = 1$ and β according to Eq. (9.7)). The magnitude of α must not exceed $2/\Lambda$ much, due to the various approximation errors inherent to the numerical evaluation of Eq. (9.6). Hence, the net acceleration of the PM by an exponential spectral transformation is strongly limited and, in particular, less than by the Chebyshev acceleration.

9.2.4 Simultaneous iteration

The vector iteration / PM is generalized to the *subspace iteration* by using M orthonormal trial vectors $\mathbf{x}^{(m,s)}$ in each iteration step in order to produce $\tilde{M} \leq M$ eigenvectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_{\tilde{M}}\}$ simultaneously, and analogous estimates of ϕ are found [277]. The continous orthonormalization prevents the $\mathbf{x}^{(m,s)}$ from converging to $\mathbf{u}_1$ altogether. The method is also known as Bauer's method or *Treppeniteration* and may be viewed as a reduced variant of the iterative extraction-orthogonalization method (IEOM) [281] applied to stationary problems. It can be speeded up considerably if in each step the $\mathbf{x}^{(m,s)}$ are replaced with *optimal approximations* $\mathbf{v}^{(m,s)}$ of $\mathbf{u}_1, \ldots, \mathbf{u}_M$ within the iterative subspace span $[(\mathbf{x}^{(m,s)})]$.

To define this optimum, there are two criteria to choose from: either the minimal Rayleigh quotient or the minimal residual norm, as mentioned above. For subspace dimensions $< N$ both criteria yield different approximations [274], and the existence of this alternative is crucial for the present development of an improved algorithm. The *Rayleigh-Ritz procedure* (RR) [274] employs the minimal Rayleigh quotient to obtain optimal eigenvector approximations within a given subspace. Alternatively, the *residual minimization method* (RM) [278], also called *variance minimisation* method [245], minimizes the residual norm. Both methods serve to extract the present, but implicit, information on an eigenvector from a subspace, i. e. their task is the *use* of information. In the following sections we focus on methods for *gaining* and *storing* information, and for the sake of simplicity we ignore the permanently existing alternative for the way to use information. The RM, its applications, and its relationship to the RR will be considered in Section 9.2.15. Until then, we only consider the RR which tacitly always may be replaced with the RM.

Applying the RR to the $\mathbf{x}^{(m,s)}$ in each step of a subspace iteration, and choosing $M \gg \tilde{M}$, defines the *simultaneous iteration* [274, 276], also called subspace iteration with projection [277]. It is superior to the subspace iteration since it makes more use of information. However, this is not yet sufficient to solve the problem of slow convergence caused by the trade-off $\phi < C/\Lambda$ with C being in the order of $\lambda_M - \lambda_1$.

9.2.5 Krylov methods

To overcome the antiproportionality between convergence rate ϕ and spread Λ it is necessary to *accumulate* the information obtained during the iteration. Complete storage in connection with extensive use of information is achieved by the following procedure due to Arnoldi [277] (we consider the calculation of $\mathbf{u}_1$ only, thus $M = 1$) which, however, is of explanatory interest only. In practice, this algorithm is not implementated in the described way for symmetric eigenproblems, for reasons explained in Section 9.2.14.

0.) Suppose that the sequence of all previous trial vectors $\mathbf{x}^{(s)}$, $s \leq S$, is stored to serve as a basis of $\mathcal{K}^{(S)} = \mathrm{span}\,[(\mathbf{x}^{(s)})]$.

1.) The best approximation of $\mathbf{u}_1$ within $\mathcal{K}^{(S)}$, the Ritz vector $\mathbf{v}^{(S)}$, is calculated by means of the RR (*use of the entire information*).

2.) Next, $\mathbf{y} = H\mathbf{v}^{(S)}$ is calculated (*gain of "new information"*).
Note that if $\mathbf{v}^{(S)} = \mathbf{u}_1$ then $\mathbf{y} = \lambda_1 \mathbf{v}^{(S)}$ will give no "new information".

3.) If so, the procedure terminates. Otherwise we orthogonalize $\mathbf{y}$ against $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}$, to eliminate any redundance from $\mathbf{y}$, obtaining a vector $\mathbf{b}$. As the approximations due to $\mathbf{v}^{(s)}$ become better, $||\mathbf{b}|| \to 0$.

4.) Finally, $\mathbf{b}$ is normalized and appended to the sequence as new trial vector $\mathbf{x}^{(S+1)}$ (*cumulative, redundance-free storage of the new information*), $S \mapsto S + 1$.

The $\mathbf{x}^{(s)}$ are orthonormal in each step S and the representation of H with respect to them is tridiagonal because H is Hermitian.

There is a relationship between this procedure and the PM. Given that none of the $\mathbf{x}^{(s)}$ is orthogonal to $\mathbf{u}_1$ the s-th trial vector is obtained from the start vector $\mathbf{x}^{(1)}$ by a polynomial X_s of H of degree $s - 1$:

$$\mathbf{x}^{(s)} = [X_s(H)]\,\mathbf{x}^{(1)} \quad \text{with} \quad X_s(H) = \sum_{n=0}^{s-1} p_{n,s} H^n \quad . \tag{9.8}$$

Consequently, $H^{s-1}\mathbf{x}^{(1)}$ is contained in $\mathcal{K}^{(s)}$ for all $s \leq S$ so that

$$\mathcal{K}^{(S)} = \mathrm{span}\,(\{\mathbf{x}^{(1)}, H\mathbf{x}^{(1)}, \ldots, H^{S-1}\mathbf{x}^{(1)}\}) \quad . \tag{9.9}$$

It follows that [277]

(i) $\mathcal{K}^{(S)}$ is *the span of all iterates of the PM* for $s \leq S$. It is called the $(S-1)$-th *Krylov space* of $\mathbf{x}^{(1)}$.

(ii) The Arnoldi procedure constructs the sequence of successive Krylov spaces $\mathcal{K}^{(s)}$ which is called the *Krylov sequence* of $\mathbf{x}^{(1)}$.

(iii) Upon each element $\mathcal{K}^{(s)}$ of this sequence, i. e. after each $H\mathbf{x}$ multiplication, the procedure executes the RR to obtain an optimal eigenvector approximation $\mathbf{v}^{(s)}$.

In summary, the Arnoldi procedure performs the PM but does not wait for $H^s\mathbf{x}^{(1)}$ to converge to $\mathbf{u}_1$. Instead, after each step it gathers the complete information, which it has stored for this purpose, and subjects it to a comprehensive evaluation (the RR) yielding $\mathbf{v}^{(s)}$ as the actual result. The resulting convergence rate of this procedure is

$$\phi \;\geq\; 2\sqrt{\gamma} \tag{9.10}$$

[274,277] (an intuitive argument for Eq. (9.10) obviating the Kaniel-Saad theory is given in Section 9.2.8). Since $\sqrt{\gamma} \gg \gamma$ (for $\gamma \ll 1$) the accumulation of information, permanently followed by the full exhaustion of the Krylov sequence generated by the start vector, introduces a substantial acceleration, as compared to the simultaneous iteration.

9.2.6 Three-term recurrences

The above described Arnoldi procedure is a multi-term recurrence explicitly using all trial vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S-1)}$ to construct the new trial vector $\mathbf{x}^{(S)}$, thus being an S-term recurrence. However, assuming exact arithmetic it is possible to perform the same algorithm by means of a three-term recurrence, $\mathbf{t}^{(s+1)} = f(\mathbf{t}^{(s)}, \mathbf{t}^{(s-1)})$, which can even be done in two different ways. The first variant constructs the orthogonal sequence $\mathbf{t}^{(s)} = \mathbf{x}^{(s)}$ and is known as *Lanczos algorithm* (LA) [274–276,282,283]. The second variant calculates the non-orthogonal sequence $\mathbf{t}^{(s)} = \mathbf{v}^{(s)}$ and is the Fletcher-Reeves type conjugate-gradient (CG) method applied to the Rayleigh quotient ρ_H [180, 270, 271, 279, 284–286]. The successive $\mathbf{v}^{(s)}$ are mutually conjugate, i. e., orthogonal with respect to the metric induced by H.

While the Arnoldi S-term recurrence runs stably as well in finite-precision arithmetic the two three-term variants do not, and their mathematical equivalence applies only for small step counts S or exact arithmetic. In finite precision arithmetic the respective trial vectors $\mathbf{t}^{(s)}$ step by step loose their orthogonality in the respective metric. The behavior is known as self-termination property of the CG [275,287] or, in the LA, as Lanczos phenomenon [274,275,277]. It reduces the efficiency because more $H\mathbf{x}$ multiplications become necessary in practice than required in theory. Measures to face this were proposed for both variants, namely restarting (truncation) for both CG [180, 285] and LA [283], and reorthogonalization for the LA [274, 276, 282].

In exact arithmetic the Fletcher-Reeves implementation of the CG is identical to the Polak-Ribiere implementation [180] since ρ_H is a quadratic form. At finite precision, however, the Polak-Ribiere algorithm switches smoothly to a gradient method as soon as self-termination occurs, thus arranging a secret restart. Therefore, it is commonly prefered over the Fletcher-Reeves algorithm. In the following we will not distinguish those variants and, instead, assume the even more favorable, ideal behavior of the CG.

The reorthogonalization which can be included into the LA effectively converts the three-term recurrence back to an $\tilde{S}$-term recurrence ($3 \leq \tilde{S} \leq S$) so that the Arnoldi S-term recurrence is equivalent to a LA with *full* reorthogonalization [288] (for Hermitian H). According to our criterion for efficiency, algorithms forcing orthogonality are preferable over approaches completely waiving reorthogonalization (as the LA variants proposed in, e. g., Ref. [275, 283] or the CG method).

There is another practical three-term recurrence which is not known to suffer from stability problems: The Chebyshev iteration (CI) [277, 289] explained in Section 9.2.8. It is analogous to the identically named algorithm for the solution of linear equation systems [287] but different from the Chebyshev accelerated PM in Section 9.2.3. The CI is closely related to the molecular dynamics method using a second order equation of motion, as detailed in Section 9.2.9.

We continue by deriving the CI from the PM. For the sake of clarity we will drop any intermediate normalization of trial vectors. Hence, a fix vector of a recurrence is no more recognized by $\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)}$ but by

$$\mathbf{x}^{(s+1)} = a \cdot \mathbf{x}^{(s)} \tag{9.11}$$

with an appropriate scalar a.

9.2.7 Convergence rate of the power method

By the two-term recurrence $\mathbf{x}^{(s+1)} = (\alpha H + \beta I)\mathbf{x}^{(s)}$ the PM constructs the image created from the start vector $\mathbf{x}^{(1)}$ by the operator polynomial

$$\mathbf{x}^{(s)} \;=\; [(\alpha H + \beta I)^{s-1}]\,\mathbf{x}^{(1)} \;=\; \sum_{j=1}^{N} (\alpha\lambda_j + \beta)^{s-1} c_j^{(1)} \mathbf{u}_j \quad , \tag{9.12}$$

reaching the fastest convergence towards the eigenvector $\mathbf{u}_1$ if, e. g.,

$$\alpha = 2/(\lambda_2 - \lambda_N) \qquad \text{and} \qquad \beta = (\lambda_N + \lambda_2)/(\lambda_N - \lambda_2) \quad . \tag{9.13}$$

Then the polynomial Eq. (9.12) maps the eigenvalues of H according to

$$\begin{aligned} \lambda_1 &\;\mapsto\; \nu_1^{(s)} = [(1+\gamma)/(1-\gamma)]^{s-1} \\ \lambda_2 &\;\mapsto\; \nu_2^{(s)} = 1 \\ \lambda_N &\;\mapsto\; \nu_N^{(s)} = (-1)^{s-1} \end{aligned}$$

onto the amplification factors of the respective eigencomponents $c_j\mathbf{u}_j$ of $\mathbf{x}^{(1)}$, see Fig. 9.1. Thus, as s increases, $\nu_3, \ldots, \nu_{N-1}$ will vanish, ν_2 and ν_N stagnate, and only ν_1 will increase. The convergence rate can be estimated from the $\nu_j^{(s)}$ by

$$\phi \;\approx\; \frac{1}{z} \lim_{s\to\infty} \left(\ln \left| \frac{\nu_1^{(s+1)}}{\nu_1^{(s)}} \right| \;-\; \frac{1}{2} \ln \frac{\sum_{j=2}^{N} |\nu_j^{(s+1)} c_j^{(1)}|^2}{\sum_{j=2}^{N} |\nu_j^{(s)} c_j^{(1)}|^2} \right) \tag{9.14}$$

provided that $||\delta\mathbf{x}^{(s)}|| \ll ||\mathbf{x}^{(s)}||$. For α and β according to Eq. (9.13) and large s both sums in Eq. (9.14) may be replaced with $|c_2^{(1)}|^2 + |c_N^{(1)}|^2$ so that

$$\phi \;\approx\; \ln\left[(1+\gamma)/(1-\gamma)\right] \;=\; 2\gamma + \mathcal{O}(\gamma^2) \quad ,$$

proving Eq. (9.2).

9.2.8 Chebyshev iteration

Instead of Eq. (9.12) the image of $\mathbf{x}^{(1)}$ created by a different operator polynomial,

$$\mathbf{x}^{(s)} \;=\; [T_{s-1}(\alpha H + \beta I)]\,\mathbf{x}^{(1)} \qquad , \tag{9.15}$$

can be calculated by the three-term recurrence

$$\mathbf{x}^{(s+1)} \;=\; 2\,(\alpha H + \beta I)\,\mathbf{x}^{(s)} \;-\; \mathbf{x}^{(s-1)} \tag{9.16}$$

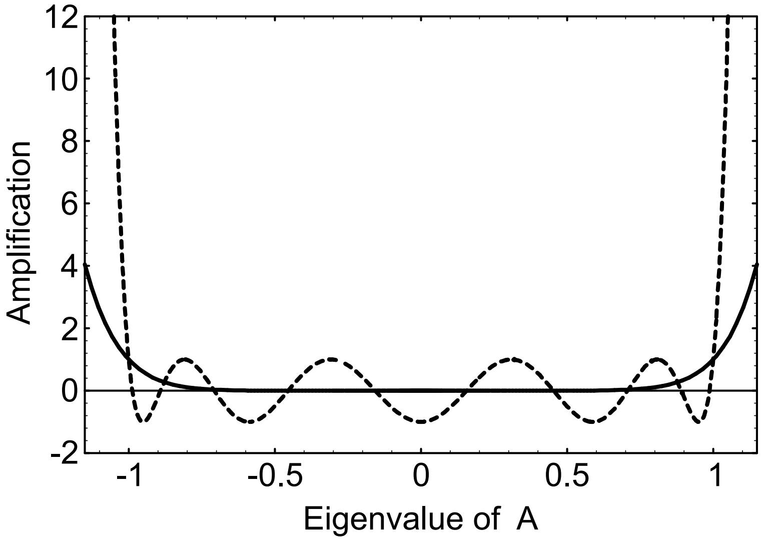


Figure 9.1:

Amplification factors $\nu^{(s)}$ versus the eigenvalues $\nu^{(2)}$ of $A = \alpha H + \beta I$ after 10 steps ($s = 11$) of the power method (solid) and of the Chebyshev iteration (dashed). For both algorithms it is optimal to choose α and β such that $\nu_2^{(2)} = 1$ and $\nu_N^{(2)} = -1$ or vice versa because then only $|\nu_1^{(2)}| > 1$. The importance of accurate shifting can clearly be seen.

constituing the CI. The advantage of the Chebyshev polynomials Eq. (9.15) is that for the α and β values from Eq. (9.13)

$$\nu_1^{(s)} = T_{s-1}[(1+\gamma)/(1-\gamma)] \approx \mathrm{e}^{2(s-1)\sqrt{\gamma}}/2 \qquad (s > 1/\sqrt{\gamma})$$

grows approximately exponentially as a function of s while all other ν_j are kept oscillating between ±1, see Fig. 9.1. Due to the harmonic oscillation of $\nu_2, \ldots, \nu_N$ the two sums in Eq. (9.14) cancel *in average*. Hence, the CI has an average convergence rate $0 \leq \phi \leq 2\sqrt{\gamma}$ which heavily depends on the actual values of α and β: Consider the operator $A = \alpha H + \beta I$ so that $\mathbf{x}^{(s)} = [T_{s-1}(A)]\,\mathbf{x}^{(1)}$ and the eigenvalues of A are $\nu_j^{(2)}$. Then the CI converges to $\mathbf{u}_1$ if and only if

$$|\nu_1^{(2)}| > \max(1, |\nu_2^{(2)}|, \ldots, |\nu_N^{(2)}|) \qquad \Rightarrow \quad \phi \geq 0 \qquad , \tag{9.17}$$

and it converges fastest to $\mathbf{u}_1$ if

$$(\min_{j=2}^{N} \nu_j^{(2)} = -1 \quad \wedge \quad \max_{j=2}^{N} \nu_j^{(2)} = 1) \qquad \Rightarrow \quad \phi \approx 2\sqrt{\gamma} \qquad , \tag{9.18}$$

as can be seen from Fig. 9.1. There are two different pairs α, β satisfying $|\nu_2^{(2)}| = |\nu_N^{(2)}| = 1$ and providing the fastest possible convergence:

$$\alpha = 2/(\lambda_2 - \lambda_N) \quad , \quad \beta = (\lambda_N + \lambda_2)/(\lambda_N - \lambda_2) \quad \text{or} \tag{9.19}$$

$$\alpha = 2/(\lambda_N - \lambda_2) \quad , \quad \beta = (\lambda_2 + \lambda_N)/(\lambda_2 - \lambda_N) \quad . \tag{9.20}$$

In practice, it needs a true fine tuning to adjust either variant, but it is possible to determine α and β automatically [289].

It is worth noticing that no other polynomial $p_k(H)$ of degree k could provide a faster increase of $\nu_1^{(k+1)}$ than $T_k(A)$, because of the min-max property of the Chebyshev polynomials [274, 287] already quoted in Section 9.2.3. Hence, in the CI not a single $H\mathbf{x}$ product could be saved by choosing A an integer rational function of degree $z > 1$: The resulting polynomial $p_{s\cdot z}(H)$ of degree $s \cdot z$ would give no larger $\nu_1^{(s\cdot z+1)}$ than $T_{s\cdot z}(A)$. In this sense Eq. (9.16) is an optimal recurrence for eigenvector calculation, if α and β are properly adjusted.

The CI is substantially superior to the PM due to the exponential growth of ν_1 , but less efficient than the LA: First, the best polynomial to calculate $\mathbf{x}^{(s)}$ from $\mathbf{x}^{(1)}$ is the one mapping $\mathbf{v}^{(1)}$ to $\mathbf{v}^{(s)}$ in the Arnoldi procedure discussed in Section 9.2.5, V_s say , which certainly is different from Eq. (9.15). Second, the performance of the LA does not depend on the choice of α and β, so the LA is run with $A = H$.

The CI is related to the PM in the sense that both methods construct a polynomial of H *independent* of the start vector $\mathbf{x}^{(1)}$. At the same time, this is the fundamental difference between the CI and the LA, reflecting a lacking use of available information. Hence the convergence rate of the LA must be at least the maximum of Eq. (9.18), as stated in Section 9.2.5.

9.2.9 Second order molecular dynamics

The molecular dynamics method relies on the numerical integration of an initial value problem of the second order equation of motion (EOM)

$$m \frac{d^2\mathbf{x}(t)}{dt^2} = -[H - \sigma(t)I]\,\mathbf{x}(t)$$

for the purpose of eigenvector computation [270] (the inclusion of damping is discussed in Section 9.2.10). A simultaneous calculation of $M > 1$ eigenvectors from the EOM necessitates the imposal of constraints to force (i) orthogonality and (ii) normalization of the trial vectors $\mathbf{x}^{(\{1,\dots,M\},s)}$ in each step s. While (i) is required to preserve prevailing information, (ii) is due to computer arithmetics only. Formally, either constraint may be added to the EOM as a set of Lagrange multipliers [270]. Alternatively, both constraints can be forced by separate orthonormalization between two iteration steps.

In the following we consider the calculation of $\mathbf{u}_1$ only, leaving (i) up to Section 9.2.10. To integrate the initial value problem numerically it is convenient to employ *Stoermer's rule* [180] (also called *Verlet algorithm* [290])

$$\mathbf{x}^{(s+1)} = 2\mathbf{x}^{(s)} - \mathbf{x}^{(s-1)} - \frac{\delta t^2}{m}(H - \sigma^{(s)} I)\mathbf{x}^{(s)} \tag{9.21}$$

obtained from the central difference discretization on the time axis with step length δt. The local discretization error of Eq. (9.21) is $\mathcal{O}(\delta t^4)$. Eq. (9.21) is formally identical with an unpreconditioned second order (or two-step) iterative method [287] for solving the homogeneous linear equation system $(H - \sigma I)\mathbf{x} = \mathbf{0}$. Thus one might expect that it is necessary to choose $\sigma = \lambda_1$ in order to obtain $\mathbf{x}^{(\infty)} = \mathbf{u}_1$. However, this is not the case: Every eigenvector $\mathbf{u}_j$ of H is a fix vector of Eq. (9.21) according to Eq. (9.11), regardless of σ. Thus, σ does *not* depend on the wanted eigenvalue.

Nevertheless, the conventional mode to evaluate Eq. (9.21) is (i) to obey the stability condition for numerical integration, $\delta t^2 \leq 4m/\Lambda$, and (ii) to choose

$$\sigma^{(s)} = \rho_H[\mathbf{x}^{(s)}] \tag{9.22}$$

[270]. Both is actually unnecessary: (i) A numerically stable integration is required to calculate a proper trajectory but not for obtaining its stationary asymptote, which is wanted here. Point (ii) actually intends to conserve the norm of the wave function [270] which is not necessary either since normalization can be forced outside of Eq. (9.21). Furthermore, (ii) causes a self-termination of the convergence which could be avoided, thus significantly increasing the performance during the final phase of convergence.

To explain how this can be achieved, we note that Eq. (9.21) is identical to Eq. (9.16) due to the relations

$$\delta t^2/m = -2\alpha \qquad \text{and} \qquad \sigma = (1-\beta)/\alpha \qquad . \tag{9.23}$$

The following analysis of the second order molecular dynamics method is based on this observation and on implicitly treating a non-stationary iteration as a composite of successive stationary ones [287]. Hence we can apply knowledge about the CI to Eq. (9.21).

For instance, we know in advance both the largest possible convergence rate, $\phi \approx 2\sqrt{\gamma}$, and how to realize it. From Eq. (9.19) and Eq. (9.23) we obtain as optimal parameters

$$\delta t^2/m = 4/(\lambda_N - \lambda_2) \qquad \text{and} \qquad \sigma = \lambda_2$$

which violate the stability condition for the numerical integration, though looking similar to the conventional choice for using Eq. (9.21). In fact, the slightly different choice of $\delta t^2/m$ employed here yields an increase of ϕ by $\gamma^{3/2}$ only, which may be considered negligible. The major advantage of the optimal CI parameters over the conventional Verlet algorithm is due to the slightly different choice of σ: Consider the conventional iteration and its asymptotic shift $\sigma^{(s)} = \lambda_1$ being the *limes* of $\rho_H[\mathbf{x}^{(s)}]$ as $\mathbf{x}^{(s)} \to \mathbf{u}_1$. In this situation $\nu_N^{(2)} = -1$ and $\nu_1^{(2)} = 1$, i. e., the condition Eq. (9.17) is violated. Consequently, as $\mathbf{x}^{(s)}$ converges to $\mathbf{u}_1$ in the conventional iteration $\sigma^{(s)}$ converges to λ_1 , and for $\sigma^{(s)} < \lambda_2$ the iteration will further and further depart from Eq. (9.18), asymptotically even violating Eq. (9.17). This results in the sub-linear, fading convergence (self-termination).

The serious draw-back of the conventional operation of Eq. (9.21) is only due to the unnecessary incorporation of the normalization of the wavefunction into the actual EOM leading to the particularly disadvantageous shifting strategy. The conventional algorithm only performs well as long as $\rho_H[\mathbf{x}^{(s)}] > \lambda_2$, i. e., while considerations of the asymptotic behavior do not yet apply. However, for the final phase the parameters should be adjusted as suggested above, in order to avoid the self-termination.

As the second possible, optimal parameter choice Eq. (9.20) and Eq. (9.23) yield

$$\delta t^2/m \;=\; -4/(\lambda_N - \lambda_2) \qquad \text{and} \qquad \sigma \;=\; \lambda_N$$

which cause a violation of the stability condition for numerical integration, as well. Also the above mentioned interpretation as solving the linear system $(H - \lambda_N I)\mathbf{x} = \mathbf{0}$ is misleading here: The optimal parameters yield an intentionally "divergent" iteration which is exclusively appropriate for eigenvector calculation. The corresponding algorithm, overcoming the self-termination problem like our previous parameter choice, was proposed in 1992 [291] and its equivalence to the CI was observed immediately thereafter [292].

Finally it follows, that no higher-order difference quotients or advanced stepping procedures for the EOM integration would be able to accelerate the local convergence further, as they eventually only yield different polynomials $p_s(H)$ which cannot improve the convergence beyond the optimality of the CI, as explained in Section 9.2.8.

9.2.10 Modified second order molecular dynamics

Previous effort to improve the conventional algorithm focused on a prolongation of δt by simplified analytic integration of the EOM [270, 293] which leads to the following recurrence: D is the diagonal of H, and $\Omega = [(D - \sigma^{(s)} I)/m]^{1/2}$, then

$$\begin{aligned} \mathbf{x}^{(s+1)} \;=\; & \;2\,[\cos(\delta t\,\Omega)]\,\mathbf{x}^{(s)} \\ & - 2\,[I - \cos(\delta t\,\Omega)](D - \sigma^{(s)} I)^{-1}(H - D)\,\mathbf{x}^{(s)} \\ & - \mathbf{x}^{(s-1)} \end{aligned} \tag{9.24}$$

(still without damping). Expanding Eq. (9.24) to the second order of δt returns Eq. (9.21), i. e., for small $\delta t^2/m$ no difference will occur in practice. To demonstrate the possible advantage of Eq. (9.24) over Eq. (9.21) for larger time steps we rewrite it in the scheme of Eq. (9.16), $\mathbf{x}^{(s+1)} = 2A\mathbf{x}^{(s)} - \mathbf{x}^{(s-1)}$. Then

$$A \;=\; I + [\cos(\delta t\,\Omega) - I](D - \sigma^{(s)} I)^{-1}(H - \sigma^{(s)} I) \tag{9.25}$$

so that $\mathbf{x}^{(s)} = [T_{s-1}(A)]\,\mathbf{x}^{(1)}$ if $\sigma^{(i)} = \sigma$ is constant for $i < s$.

Thus the CI is no longer performed with a linear transform $A(H)$ but with a non-analytic function $A(H)$ which does not commute with H. Generally, such a procedure is only appropriate if the desired, corresponding eigenvectors of H and A do not differ too strongly and if the related spectral transformation does not coalesce distinct eigenvalues of H within the region of interest. To examine under which condition this is the case we consider first a diagonal H, thus $D = H$. Then, from Eq. (9.25) follows

$$\begin{aligned} A \;&=\; \cos\,[(H - \sigma I)\delta t^2/m]^{1/2} \qquad \text{and} \\ \nu_j^{(s)} \;&=\; T_{s-1}\left(\cos\,[(\lambda_j - \sigma)\delta t^2/m]^{1/2}\right) \\ &=\; \begin{cases} \cosh\left[(s-1)\sqrt{(\sigma - \lambda_j)\delta t^2/m}\right] & \text{if} \quad \lambda_j < \sigma \\ \cos\left[(s-1)\sqrt{(\lambda_j - \sigma)\delta t^2/m}\right] & \text{if} \quad \lambda_j \geq \sigma \end{cases} . \end{aligned}$$

These expressions are the exact solution of the EOM with respect to $c_j^{(s)}/c_j^{(1)}$ at time $t = (s-1)\delta t$ so that the modified CI performs an exact integration of the EOM when H is diagonal. Also here, Eq. (9.18) requires $\sigma = \lambda_2$. Then Eq. (9.14) yields $\phi \approx \sqrt{(\lambda_2 - \lambda_1)\delta t^2/m}$. This is an improvement over Eq. (9.15) whenever $\delta t^2/m > 4/\Lambda$, i. e., for larger time steps than "allowed" for the conventional integration: For diagonal H the properly shifted cosine transformation Eq. (9.25) enables an *arbitrary* acceleration without any deflection of eigenvectors. This improvement is, however, conditional on keeping $\sigma^{(s)} \geq \lambda_2$ or, at least, $\sigma^{(s)} > \lambda_1$: Permitting $\sigma^{(s)} \to \lambda_1$ by the use of Eq. (9.22) would cause $\phi \propto \sqrt{\sigma^{(s)} - \lambda_1} \to 0$ in the advanced procedure, as well, and would – regardless of the time step length – cause a self-termination due to the asymptotic violation of Eq. (9.17).

For non-diagonal H, $H \neq D$ and A from Eq. (9.25) is not necessarily Hermitian anymore and $[A, H] \neq 0$, so that in the general case the replacement of H by A

in the CI will *deflect* the eigenvectors, i. e. the stationary solutions of the EOM, and introduce a systematic error $\mathbf{x}^{(\infty)} \neq \mathbf{u}_1$ superponing the discretization error introduced by the basis $\mathcal{B}$. There are, however, exceptions: $\mathbf{u}_j$ is an eigenvector also of $\lim_{\sigma \to \lambda_j}(A)$ (to the eigenvalue 1), since in the limit the last term in Eq. (9.25) will vanish by multiplication with $\mathbf{u}_j$. It follows that the use of A according to Eq. (9.25) is *consistent* [287] with the original eigenproblem $H\mathbf{x} = \lambda_1 \mathbf{x}$ if and only if $\sigma^{(s)} \to \lambda_1$ for $s \to \infty$. Hence only the conventional approach using Eq. (9.22) exhibits no deflection of $\mathbf{x}^{(\infty)}$ and gives the exact eigenvector $\mathbf{u}_1$ for a general (non-diagonal) Hermitian H.

By comparing Eq. (9.25) with the preconditioned second order iterative method for solving $(H - \lambda_1 I)\mathbf{x} = \mathbf{0}$ [287] we identify the preconditioning matrix

$$C \;=\; [I - \cos(\delta t\, \Omega)]^{-1}(D - \sigma^{(s)} I) \qquad . \tag{9.26}$$

Hence the algorithm proposed in Ref. [293], based on analytic integration of the EOM, may be viewed as a *diagonally preconditioned CI*. The preconditioning by Eq. (9.26) and Eq. (9.22) may enhance the convergence even for dense H but the acceleration is particularly significant for diagonally dominant H, as demonstrated in Ref. [293]. However, forced by the demand for consistency, the convergence rate decreases when $\sigma^{(s)}$ becomes smaller then λ_2, which is the same situation as for diagonal H.

The escape to nevertheless keep $\sigma^{(s)} \geq \lambda_2$ implies eigenvector deflection, $\theta = \angle\,(\mathbf{x}^{(\infty)}, \mathbf{u}_1) \neq 0$, so that the use of A effects a substitution of H rather than a preconditioning. One might accept this if $||\delta\mathbf{x}^{(\infty)}|| = \sqrt{2[1 - \mathrm{Re}(\cos\theta)]}$ could be kept in the order of magnitude of the discretization error. In turn, this might require to fulfill the stability condition $\delta t^2/m \leq 4/\Lambda$ so that, eventually, the use of Eq. (9.25) would be less efficient than the linear transformation Eq. (9.19).

From this dilemma we conclude that the preconditioned CI using Eq. (9.26) and Eq. (9.22) may, depending on H, be advantageous as long as $\sigma^{(s)} \geq \lambda_2$, but should be replaced with the normal CI using Eq. (9.19) in the final phase of convergence where $\rho_H[\mathbf{x}^{(s)}] < \lambda_2$.

Damping

In the case of diagonally dominant H the performance of the molecular dynamics method can be improved by introducing damping to the EOM [270]. This leads to a different recurrence,

$$\mathbf{x}^{(s+1)} \;=\; 2EB\mathbf{x}^{(s)} \;-\; E^2\mathbf{x}^{(s-1)} \qquad ,$$

which, for constant σ, is equivalent to

$$\begin{aligned}
\mathbf{x}^{(s)} &= [E^{s-1}T_{s-1}(B)]\,\mathbf{x}^{(1)} \qquad \text{with} && (9.27)\\
B &= [\cosh(\eta\,\delta t\,|\Omega|)] + [\cos(\delta t\,\Omega) - \cosh(\eta\,\delta t\,|\Omega|)](D-\sigma I)^{-1}(H-\sigma I) && (9.28)\\
E &= [\exp(-\eta\,\delta t\,|\Omega|)] && (9.29)\\
|\Omega| &= [|D-\sigma I|/m]^{1/2} &&\\
\eta &\geq 0 \qquad . &&
\end{aligned}$$

For diagonal H we obtain $B = A$ and $E = \exp[-\eta\,(|H-\sigma I|\,\delta t^2/m)^{1/2}]$ and

$$\nu_j^{(s)} = \begin{cases} \exp[-\eta(s-1)\sqrt{(\sigma-\lambda_j)\delta t^2/m}]\ \cosh\left[(s-1)\sqrt{(\sigma-\lambda_j)\delta t^2/m}\right] \\ \quad \text{if} \quad \lambda_j < \sigma\,, \\ \exp[-\eta(s-1)\sqrt{(\lambda_j-\sigma)\delta t^2/m}]\ \cos\left[(s-1)\sqrt{(\lambda_j-\sigma)\delta t^2/m}\right] \\ \quad \text{if} \quad \lambda_j \geq \sigma \end{cases}$$

in the same manner as above. Eq. (9.14) with $\lambda_1 < \sigma < \lambda_2$ gives

$$\phi \approx \frac{\delta t}{\sqrt{m}}\left[(1-\eta)\sqrt{\sigma-\lambda_1} + \eta\sqrt{\lambda_2-\sigma}\right] \qquad .$$

It follows that $\eta = 1$ is the optimal damping if $\sigma < (\lambda_1+\lambda_2)/2$ and $\eta = 0$ else. Thus the asymptotic convergence rate for $\sigma \to \lambda_1$ is *not* zero anymore but $\phi \approx \sqrt{(\lambda_2-\lambda_1)\delta t^2/m}$, i. e., the conventional use of Eq. (9.22) causes no self-termination anymore. Hence, for diagonal H the damping offers a great advantage over both Eq. (9.19) and Eq. (9.26): There is arbitrary acceleration of the convergence even for $\rho_H[\mathbf{x}^{(s)}] < \lambda_2$.

However, for non-diagonal H ($D \neq H$) Eq. (9.27) is *never* consistent with $H\mathbf{x} = \lambda_1\mathbf{x}$, except for $\eta = 0$. Since $\eta = 0$ is equivalent to Eq. (9.26) we conclude that damping is *not* useful for Hamiltonians which are not diagonally dominant.

The preconditioning of iteration Eq. (9.21) by Eq. (9.26) succeeded to reduce the required number of $H\mathbf{x}$ products by up to 90% for a diagonally dominant H [270]. For not diagonally dominant H, as the discrete eight-band **kp** Hamiltonian, such savings should not be expected.

The dilemma between sublinear asymptotic convergence and inconsistency, arising for general choices of H, is not solved by analytical integration of the EOM or damping, but only by implementations of Eq. (9.21) using $\delta t^2/m = 4/(\lambda_N - \lambda_2)$ and a fixed energy shift $\sigma^{(s)} = \lambda_2$ if $\rho_H[\mathbf{x}^{(s)}] < \lambda_2$. Thus the molecular dynamics method merges into the CI, and the CI with automatical adjustment of α, β as proposed in Ref. [289] would be the fastest numerical integration of the Newtonian second order EOM. To compute more eigenvectors the serial procedure with implicit deflation [274, 277, 289, 291, 294] could be used.

We finish the discussion of the second order molecular dynamics method with the conclusion that none of the proposed variants gives a better performance than the CI in the general case, so that the capability of the method ranks below that of the LA, as regards the solution of large Hermitian eigenproblems.

9.2.11 Modified first order molecular dynamics

Analogous considerations apply to the molecular dynamics method using a first order EOM [270],

$$m\,\frac{d\mathbf{x}(t)}{dt} \;=\; -[H-\sigma(t)I]\,\mathbf{x}(t) \qquad ,$$

to be integrated by the explicit Euler method,

$$\mathbf{x}^{(s+1)} \;=\; \mathbf{x}^{(s)} - \frac{\delta t}{m}(H-\sigma^{(s)}I)\mathbf{x}^{(s)} \qquad . \tag{9.30}$$

Again, the constraints can be imposed separately and need not be incorporated into the actual EOM. Eq. (9.30) is formally identical with an unpreconditioned first order (or one-step) iterative method [287] for solving $(H-\sigma I)\mathbf{x}=\mathbf{0}$. Further, Eq. (9.30) is a gradient method (or steepest-descent method) to minimize the Rayleigh quotient ρ_H [279]. Further, it is equivalent to the PM Eq. (9.12) by $\delta t/m = -\alpha$ and $\sigma = (1-\beta)/\alpha$. Hence, we know in advance that the fastest possible convergence to $\mathbf{u}_1$ (with $\phi \approx 2\gamma$) is obtained whenever

$$\frac{\delta t}{m} \;=\; \frac{2}{\lambda_2+\lambda_N-2\sigma} \qquad .$$

Like Eq. (9.21) the iteration Eq. (9.30) can be speeded up by analytic integration of the EOM. This leads to the PM, $\mathbf{x}^{(s+1)} \;=\; A\mathbf{x}^{(s)}$, applied to

$$A \;=\; I + (\exp[-(D-\sigma^{(s)}I)\delta t/m] - I)(D-\sigma^{(s)}I)^{-1}(H-\sigma^{(s)}I) \tag{9.31}$$

[270]. From this we identify the preconditioning matrix

$$C \;=\; (I-\exp[-(D-\sigma^{(s)}I)\delta t/m])^{-1}(D-\sigma^{(s)}I) \tag{9.32}$$

so that the algorithm using A is a preconditioned gradient method. Expanding Eq. (9.31) to first order of δt returns Eq. (9.30) which is consistent with $H\mathbf{x}=\lambda_1\mathbf{x}$ if and only if $\sigma=\lambda_1$. For diagonal H we find

$$\nu_j^{(s)} \;=\; \exp\left[-\frac{\delta t}{m}(s-1)(\lambda_j-\sigma)\right] \qquad .$$

Applying Eq. (9.14) with $\sigma = \lambda_1$ we obtain $\phi \approx (\lambda_2 - \lambda_1)\delta t/m$ which is an improvement over the unpreconditioned variant if $\delta t/m > 1/\Lambda$.

The next question is how the first order EOM compares to the second order EOM in case of non-diagonal H.

9.2.12 Comparison of methods

The efficiency of a given method is determined by three independent criteria:

(i) The basic algorithm: Is it equivalent to the PM, the CI, the CG, the LA, or etc.?

(ii) The preconditioner: How well is it suited for the given H?

(iii) The preconditioner: How well is it suited for the basic algorithm? In particular, does it support the optimal operating conditions for the basic algorithm?

Any comparison of methods should take into account these three questions separately.

Concerning (i) we summarize the results from the preceeding sections: The least efficient basic algorithm to calculate the "groundstate" $(\tilde{\mathbf{u}}_1, \tilde{\lambda}_1)$ of an operator A is the PM. It constructs the most primitive sequence of polynomials of A, (A^s), in order to approximate the projector Π_1 into the eigenspace spanned by $\tilde{\mathbf{u}}_1$. If $\tilde{\gamma}$ denotes the gap ratio of A assigned to $\tilde{\lambda}_1$ the optimally shifted PM has the effective asymptotic convergence rate $\phi \approx 2\tilde{\gamma}$. The PM is equivalent to gradient algorithms and to the 1st order EOM method. Better by far is the CI constructing the sequence $(T_s[A])$. This sequence enables the fastest convergence towards Π_1 possible, and the optimally shifted CI exhibits an average convergence rate of $\phi \approx 2\sqrt{\tilde{\gamma}}$.

Given a particular start vector $\mathbf{x}^{(1)}$ the actual intention behind constructing Π_1 is to delete all eigencomponents of $\mathbf{x}^{(1)}$ being orthogonal to $\tilde{\mathbf{u}}_1$. Surely, this process can be supported by granting some influence on the iteration procedure to the start vector itself, i. e., by constructing the optimal sequence $(V_s[A])$ (see p. 135) *dependent* on the particular start vector $\mathbf{x}^{(1)}$. This opportunity is ignored by both the PM and the CI, but exhausted by the Arnoldi procedure in Section 9.2.5 which ensures complete use of the available information.

The complete elimination of redundancy from the subspace basis leads to mutual orthogonality of all trial vectors $\mathbf{x}^{(s)}$. The convergence rate of the Arnoldi procedure is $\phi \geq 2\sqrt{\tilde{\gamma}}$, independent of any linear adjustment of A like that necessary in the PM and CI. There are more economical implementations of the

Arnoldi procedure for Hermitian A, maintaining the orthogonality even without performing a full orthogonalization in each step. These are the LA with partial or selective *re*orthogonalization. There are also implementations which give up the orthogonality (with respect to the metrics induced by I or A) under the impact of round-off errors, as the LA without reorthogonalization or the CG. These methods need a larger number of $A\mathbf{x}$ products than those forcing orthogonality and are therefore less efficient in our sense.

Before considering question (ii) we recall that the only way to accelerate a given basic algorithm is to modify the spectral structure of the respective operator, $H \mapsto A$, in such a way that $\gamma_A > z\gamma_H$ while $\mathbf{u}_1$ remains unaltered [277] (z is the computational expense for evaluating $A\mathbf{x}$ expressed in units of the expense for a $H\mathbf{x}$ multiplication). We may divide the possible functions $f : H \mapsto A$ to achieve this goal into two classes. The first class leaves all eigenvectors unaltered by $[A, H] = 0$ and thus consists of analytic f only. As shown in Section 9.2.3 there is no increase of γ_A beyond Eq. (9.5) by this class. The enhancement obtained thus is sufficient to accelerate the PM but not the CI: From Eq. (9.5) follows that the effective convergence rate of the CI will even *de*crease by $\mathcal{O}(\gamma)$ for $z > 1$, and the same holds for the LA. Hence, the only possibility to accelerate the CI or LA is to choose f from the second class characterized by $[A, H] \neq 0$. Next we consider questions (ii) and (iii) which refer to this second class, the non-analytic functions, only.

9.2.13 Preconditioning

The aim of preconditioning is to spread the spectrum of H around λ_1 and to compress it elsewhere in order to increase γ. The most favorable arrangement would be that all eigenvalues of A except for μ_1 are $(N-1)$-fold degenerate since then $\gamma_A = 1$. The minimal demand is that the eigenvector $\mathbf{a}_1$ of A corresponding to μ_1 is related to the eigenvector $\mathbf{u}_1$ of H corresponding to λ_1 in an exactly known and unique way. Traditionally, this is attempted by replacing H with

$$\tilde{A} = L^{-1}(H - \sigma I)(L^{\dagger})^{-1} \tag{9.33}$$

where L is a lower triangular matrix. $\tilde{A}$, L, and σ may depend on the step count s, and $\sigma \to \lambda_1$ for $s \to \infty$. $\tilde{A}$ is Hermitian like H and $\tilde{\mathbf{a}}_1 = L^{\dagger}\mathbf{u}_1$. Thus $\tilde{A}$ is a feasible substitute for H in any of the basic algorithms discussed, in particular the CG and the symmetric LA. We emphasize on these two not only because they are most efficient but also because they strongly rely on the Hermiticity of the used operator $\tilde{A}$, unlike the PM or the CI. The implicit substitution of H by $\tilde{A}$ in the CG is standard and produces the PCG [180, 270, 271, 284, 287]. The explicit substitution within the truncated, symmetric LA embedded into an outer loop for changing σ and L gives the *preconditioned LA* which has been demonstrated to be very efficient, too [295].

$\tilde{A}$ has the same spectrum as

$$A = C^{-1}(H - \sigma I) \tag{9.34}$$

where $C = LL^{\dagger}$. Again, C and σ may depend on the step count s. Clearly, the preconditioning matrix C shall be a possibly good, but first of all easily invertible, approximation of $H - \sigma I$ in order to produce the desired spectral properties of A stressed above.

At this point it is important to remember the following: If the almost singular operator $H - \sigma I$ should be entirely invertible to acceptable accuracy at moderate expense in whatsoever manner, then inverse iteration [21, 180, 274, 277, 296], Rayleigh quotient iteration [277, 297], or a shift-and-invert LA [277, 282] should be tried. The outstanding virtue of these methods is that they not only find extreme eigenpairs very fast, but also interior eigenvalues as rapidly, thus solving completely satisfactory the computational problem discussed in this chapter. However, it was our initial assumption that the shift-and-invert spectral transformation is not applicable to the Hamiltonian H.

For general C, A according to Eq. (9.34) is not Hermitian, but consistent with H if $\sigma = \lambda_1$ (then $\mathbf{a}_1 = \mathbf{u}_1$). It is thus no feasible substitute for H in the CG or the symmetric LA. However, it is possible in a different way to improve the Arnoldi procedure from Section 9.2.5 by means of a preconditioner like Eq. (9.34), leading to the GDA. For that purpose C can be generalized to indefinite matrices since a Cholesky factorization is not required anymore. This offers better opportunity for the construction of approximants to the indefinite operator $H - \sigma I$ than under the restriction to positive-definite C [298]. Therefore, A may be appropriate despite of being non-Hermitian.

If we hence waive the requirement for the existence of L, question (ii) on p. 142 reads: How to choose C for a given H? Here is where art begins. A large number of preconditioners C derived by splitting of $H - \sigma I$ has been proposed. None of them can be judged independently of the concrete Hamiltonian expansion H. We will therefore only stress a few general aspects.

A problem is that many standard preconditioning techniques [287, 299], like SOR, incomplete factorization, or SSOR, require explicit access to the matrix elements of H. This is impossible, e. g., when a mixed representation of H is used in order to avoid the handling of dense matrices like in the SM preferred for pseudopotential calculations. In our implementation of the EFA/FDM there is no access to single matrix elements either since H is encoded as a procedure applying a linear mapping to the input vector. In such cases only very few matrix elements of H are explicitly known, like the diagonal ones and some selected off-diagonal entries. Therefore, diagonal or otherwise very sparse matrices C happen to become relevant preconditioners although, e. g., the SSOR preconditioner [300] usually would give a very much larger γ_A.

Certainly, for any given $H-\sigma I$ there exists an optimal diagonal preconditioning matrix $C_{\max}$ producing the largest γ_A value possible for diagonal preconditioners. Depending on H, the corresponding $\gamma_{A,\max}$ will be significantly or only a little bit greater than γ. Since $C_{\max}$ is unknown in the general case $\gamma_{A,\max}$ will rarely be achieved, however, a suitable surrogate choice for C might yield a γ_A value close to the optimum. For instance, according to first-order perturbation theory the optimal correction vector to a trial vector $\mathbf{v}$ is obtained by multiplication with A from Eq. (9.34) and

$$C \;=\; D-\sigma I \qquad \text{and} \qquad \sigma \;=\; \rho_H(\mathbf{v}) \tag{9.35}$$

where D is the diagonal of H [298]. This preconditioner is used in the classical Jacobi method for linear equation systems [287] and known to be efficient for diagonally dominant H. Experiments with dense random matrices

$$H-\sigma I \;=\; U\,\mathrm{diag}(0,1/N,\ldots,1-1/N)\,U^\dagger$$

with unitary random U and ranks $N \le 21$ show that Eq. (9.35) yields $\gamma_A \approx \sqrt{\gamma_H\,\gamma_{A,\max}}$, thus improvement due to a diagonal preconditioner like Eq. (9.35) is not conditional on the diagonal dominance or sparsity of H [298].

The performance of a given diagonal preconditioner C must be checked in the particular case. To give suggestions, we point to Eq. (9.26) and Eq. (9.32). A diagonal preconditioner specially designed for Hamiltonians arising in pseudopotential calculations is derived as a Padé approximant to the kinetic energy operator [270]:

$$\begin{aligned} C &= (27I+18K+12K^2+8K^3)^{-1}(27I+18K+12K^2+8K^3+16K^4) \quad , \\ K &= (D-[V])/k_s \end{aligned}$$

where $[V]$ is the (diagonal) potential energy matrix and k_s is the (scalar) average kinetic energy in the state $\mathbf{x}^{(s)}$. For large wave numbers this preconditioner behaves similar to Eq. (9.35) in pseudopotential applications [272]. A related non-diagonal preconditioner is the modified Jacobi operator [301]. Whenever a non-diagonal C is at hand it should be tried [298]. In particular, block-diagonal C may enhance the performance significantly [302]. It might also be considered to evaluate Eq. (9.34) in an other representation than the usual one for H using $\mathcal{B}$ [278].

Question (iii) on p. 142 is already answered in parts: The original LA and the CG require a symmetric positive definite preconditioner in order to find a Hermitian A to operate on. Unless a shift-and-invert transformation is affordable, these algorithms will be less flexible to employ various preconditioners helpful for accessing interior eigenvalues. In contrast, the PM and CI require, in the general case, a special linear transformation in order to arrange the optimal location of the μ_j with respect to zero (in the PM) or to make them satisfy Eq. (9.17) and Eq. (9.18) (in the CI), respectively. A linear transformation does not affect γ_A, of course. However, as shown in the discussion of the molecular dynamics method both the PM and the CI are open for potentially indefinite preconditioners. The CG, LA, and the GDA (which is considered next) need no linear adjustment due to their employment of the RR or residual minimization and, unlike the CG and LA, the GDA may benefit also from indefinite preconditioners. Hence the GDA offers the greatest flexibility for different preconditioners among all methods discussed in this chapter.

9.2.14 Generalized Davidson algorithm

Preconditioning shall improve the acquisition of information. Hence, to improve the Arnoldi procedure in Section 9.2.5 it is actually sufficient to substitute H by A in step 2 only, rather than also in step 1 (see p. 130). The resulting procedure is the GDA [298, 303]. The original Davidson algorithm (DA) [304–308] is defined by Eq. (9.35) while the GDA leaves the choice of C open. Since the DA is a particular case of the GDA we consider the GDA [306, 307, 309] only.

Even for the calculation of several eigenvectors we prefer to use the single-vector variant in connection with implicit deflation [274, 277, 294]. Simultaneous variants (block DA) are outlined in [306, 307, 309–312]. The single-vector GDA for the calculation of the lowest eigenvalue and corresponding eigenvector of H, in an implementation analogous to that of the improved algorithm to be proposed below for the calculation of interior eigenpairs, proceeds as follows (S is the cycle counter starting at zero):

0.) If $S = 0$ then (*initialisation*)

   (a) an N-dimensional normalized random vector $\mathbf{x}^{(1)}$ with Gaussian deviates as components is generated and stored,

   (b) its image $\mathbf{h}^{(1)} = H\mathbf{x}^{(1)}$ is calculated and stored,

   (c) the scalar product $\mathbf{h}^{(1)^\dagger}\mathbf{x}^{(1)}$ is calculated and stored as the (1,1) matrix element of the 1×1 Ritz matrix $R^{(1)}$,

   (d) $S = 1$.

Otherwise, the two sequences of (i) the previous trial vectors $\mathbf{x}^{(s)}$ and (ii) their images $\mathbf{h}^{(s)} = H\mathbf{x}^{(s)}$ ($s = 1, \ldots, S$) are prevailing, forming the columns of the two $N \times S$ matrices $\mathbf{X}^{(S)}$ (i) and $\mathbf{H}^{(S)}$ (ii), respectively.

The $S \times S$ Ritz matrix $R^{(S)} = \mathbf{H}^{(S)\dagger}\mathbf{X}^{(S)}$ is prevailing, as well.

1.) If there exists a prevailing Ritz vector $\mathbf{v}^{(S-1)}$ from the previous cycle it is stored as a back-up potentially needed during a basis truncation in the following step.

2.) RR: The Ritz vector $\mathbf{v}^{(S)} = \sum_{s=1}^{S} a_s\mathbf{x}^{(s)}$ is calculated from the S-dimensional eigenvector a belonging to the lowest eigenvalue α of $R^{(S)}$ (*use of information*).

If storing of two further vectors ($\mathbf{x}^{(S+1)}$ and $\mathbf{h}^{(S+1)}$) would exceed the storage capacity then (*truncation*)

(a) all $\mathbf{x}^{(s)}$ and $\mathbf{h}^{(s)}$ are disposed ($s = 1, \ldots, S$),
(b) the back-up Ritz vector $\mathbf{v}^{(S-1)}$ is orthonormalized against $\mathbf{v}^{(S)}$,
(c) both vectors are stored as $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, respectively,
(d) their images $\mathbf{h}^{(s)} = H\mathbf{x}^{(s)}$ ($s = 1, 2$) are calculated[3] and stored,
(e) the 2×2 Ritz matrix $R^{(2)}$ is calculated,
(f) $S = 2$.

Otherwise the Ritz vector back-up $\mathbf{v}^{(S-1)}$ is disposed.

3.) The image $\mathbf{w} = \sum_{s=1}^{S} a_s\mathbf{h}^{(s)}$ of the Ritz vector $\mathbf{v}^{(S)}$ is calculated.

4.) The residual $\mathbf{r} = \mathbf{w} - \alpha\mathbf{v}^{(S)}$ is calculated. If $||\mathbf{r}||_2 <$ tol the GDA terminates.

5.) If a preconditioner C is at hand (which is standard for the GDA) then

(a) σ is chosen according to a shifting strategy (for instance, $\sigma = \alpha$),
(b) The pseudoresidual $\mathbf{y} = A\mathbf{v}^{(S)} = C^{-1}[\mathbf{w} - \sigma\mathbf{v}^{(S)}]$ is calculated (*gain of information*).

Otherwise, i. e. without preconditioning, $\mathbf{y} = \mathbf{w}$.

We call $\mathbf{y}$ a *correction vector* as it will serve to correct span$[\mathbf{X}^{(S)}]$ towards the wanted eigenspace span$(\mathbf{u}_1)$ of H.

---

[3] These image vectors could, instead, have been constructed from the prevailing $\mathbf{H}^{(S)}$ if the latter had not been disposed in step (2a). However, it is more opportune to calculate the new $\mathbf{H}^{(2)}$ by actual $H\mathbf{x}$ multiplications to prevent a cumulation of round-off errors during the entire GDA.

6.) The correction vector $\mathbf{y}$ is orthonormalized against the trial vector basis $\mathbf{X}^{(S)}$ yielding $\mathbf{b}$ (*elimination of redundancy*).

7.) $\mathbf{b}$ is appended to $\mathbf{X}^{(S)}$ as a new column $\mathbf{x}^{(S+1)}$ yielding $\mathbf{X}^{(S+1)}$ (*storage of information*).

8.) The image $\mathbf{h}^{(S+1)} = H\mathbf{x}^{(S+1)}$ is calculated (*gain of information*) and appended as a new column to the image vector set $\mathbf{H}^{(S)}$ yielding $\mathbf{H}^{(S+1)}$ (*storage of information*).

9.) The Ritz matrix $R^{(S)}$ is updated by adding a new column $c = \mathbf{X}^{(S+1)^\dagger}\mathbf{h}^{(S+1)}$ as well as a new row $c^\dagger$, yielding $R^{(S+1)}$.

10.) $S = S + 1$ and the next cycle starts with step 1.

In step (6) the modified Gram-Schmidt procedure [274, 294] is explicitly performed until all scalar products $|<\mathbf{b}|\mathbf{x}^{(s)}>|$ are less than the machine precision ϵ (*full* orthogonalization) . Although it has been shown [276] that *semiorthogonalization*, i. e. orthogonalization to the *half* machine precision ($\sqrt{\epsilon}$), is sufficient to produce stable and convergent algorithms, like in the LA, the full orthogonalization in our implementation is motivated by the use of *single-precision*, i. e. 4-Byte, floating point numbers: The relative precision of this number representation is $\epsilon = 1.2 \times 10^{-7}$ and may be compared to the possible rank of H, $N \approx 10^7$, in order to visualize the urgent need not to waste one digit of accuracy during as much as 1000 cycles. Accordingly, the evaluation of scalar products is performed in double precision and only the result is recast to single precision.

The reason for the preference for single-precision over double-precision floating point numbers is that thus the available memory capacity is effectively doubled, being capable of additionally storing all image vectors $\mathbf{h}^{(s)}$ without the need for earlier, more frequent truncation in step (2). In our implementation $\mathbf{X}$ and $\mathbf{H}$ are, like all other data, entirely kept in the core memory so that no read access to disks ever occurs (and write access only for the purpose of saving intermediate or the final results). Hence, periodical truncation down to two vectors of either category ($\mathbf{x}$ and $\mathbf{h}$) in step (2) is necessary. This temporarily slows down the convergence somewhat, but in the final phase the GDA never becomes slower than a CG using the same preconditioner [306]. Further economization towards less storage demand or selective reorthogonalization does not appear possible as the GDA is a true S-term recurrence.

The stopping criterion in step (4) ensures that the procedure stops if the wanted eigenvector $\mathbf{u}_1$ is found rather than any other.

In the above implementation the procedure to gain information (according to the definition on p. 126) is distributed over steps (5b) – multiplication with A –

and (8) – multiplication with H, thus interlacing with the redundancy elimination and information storage. This complication (from the explanatory viewpoint) saves one $H\mathbf{x}$ multiplication per cycle, as can be seen from a comparison with the more easily explained, basic Arnoldi procedure described in Section 9.2.5: The update of the Ritz matrix needed in step (1) of the Arnoldi procedure tacitly requires one multiplication of H with the new basis vector $\mathbf{x}^{(S+1)}$ appended in step (4), in addition to the multiplication $H\mathbf{v}^{(S)}$ in step (2). The saving of one $H\mathbf{x}$ multiplication per cycle is conditional on the possibility to store the image vectors $\mathbf{h}^{(s)}$. If this is not possible, the GDA becomes particularly expensive, as compared to the LA or CG, due to an *a priori* complexity of $z = 2$ (see p.126).

Since the GDA employs, partly, different operators to exploit information (H in step (2)) and to acquire information (A in step (5b) and H in step (8)) it is important that $\theta = \angle(\mathbf{a}_1, \mathbf{u}_1) \approx 0$, i. e. that A and H are consistent: Otherwise, step (5b) will deliver information on useless, in the worst case random-like, hyperspace directions which thus fail to establish any progress. In such a case the unsuitably chosen precondioner C prevents the, otherwise extremely robust, Arnoldi-type procedure from working properly since the correction vectors are systematically destroyed instead of improved. In the following we discuss possibilites how such situations arise and how they might be avoided.

Preconditioners for the GDA

The angle θ between the eigenvectors $\mathbf{a}_1$ of A and $\mathbf{u}_1$ of H certainly vanishes for $\sigma = \lambda_1$ in step (5a) of the GDA but, as λ_1 is unknown, σ will always take on guess values instead: σ is either chosen to be the Ritz value α from step (2), $\sigma = \rho_H[\mathbf{v}^{(S)}]$, or another currently best approximation to λ_1 [298] potentially depending on the cycle count S. We refer to the decision procedure for choosing σ as *shifting strategy*.

For practical considerations (rather than rigorous mathematical conclusions) we may define an interval $\Sigma(C) =]\sigma_{\min}; \sigma_{\max}[$ such that $\theta \approx 0$ whenever $\sigma \in \Sigma(C)$, and call $\Sigma(C)$ the *consistency range* corresponding to C. Via this definition we have introduced a relation between the actually independent components of the preconditioned operator A in Eq. (9.34), i. e., the spectral shift σ and the preconditioner C. From this follows a requirement to the preconditioner C which already was implicitly discussed in connection with the self-termination property of the modified molecular dynamics methods:

> The consistency range $\Sigma(C)$ corresponding to C should be both (i) well-positioned and (ii) as large as possible, in order to avoid a collision with the shifting strategy.

This condition for A is additional to its obligation to realize $\gamma_A > \gamma_H$. The violation of the consistency condition, $\sigma \notin \Sigma$, is a typical source of failure for the GDA, concerning the global convergence properties [298]. The Arnoldi procedure in 9.2.5, being equivalent to $C = I$, exhibits an infinite consistency range and thus tolerates potentially any shifting strategy.

A common way to construct the preconditioner C is the ansatz $C = M - \sigma I$. This ansatz requires, beside $\sigma \in \Sigma$, that σ is not too close to an eigenvalue of M: Otherwise, C is ill-conditioned [277, 287] and its numerical inversion in step (5b) will produce a random-like correction vector $\mathbf{y}$, regardless of $\sigma \in \Sigma$, due to cumulated round-off errors. This is a further source of failure in the GDA. When using diagonal M it is trivial to control [298, 307] but for different structures of M a sophisticated monitoring may be required to discover ill-conditioned C prior to their use. A common procedure to solve the problem is to change σ a little in case it happens to produce an ill-conditioned C. Non-diagonal C, such as tridiagonal [114], may be efficient but are distinctly more difficult to handle than diagonal C, due to the above outlined error sources.

Convergence rate of the GDA

It is difficult to estimate the asymptotic convergence rate ϕ of the GDA because of the characteristic task splitting between H and A. We may only give a few semi-quantitative arguments [298] which primarily are justified by observations made in practice.

If only A was used in the steps (2) and (5b), and if neither σ nor C would vary dependent on S, the GDA would behave similar to a preconditioned LA (apart from the auxiliary overhead) and construct a Krylov space generated by A. In this case

$$\phi > 2\sqrt{\gamma_A} \tag{9.36}$$

could be transfered from Eq. (9.10). The assumptions $\sigma, C = \text{const.}$ are reasonable for the local phase of convergence (which ϕ refers to) as the shifting strategy will have to adjust $\sigma \approx \lambda_1$ in the end, in order to preserve the consistency of A and H. If then in step (2) H is replaced with A, the intermediate eigenvector approximations $\mathbf{v}^{(S)}$ will change insignificantly, due to the consistency of A and H for $\sigma \to \lambda_1$, and hence the Krylov space will remain nearly unaltered.

Therefore, we assume Eq. (9.36) for the local convergence rate of the GDA. The essential content of Eq. (9.36) is that ϕ is dominated by the operator used to gain information, A in step (5b), rather than by the one used to exploit information, H in step (2). The possibility for such a task separation between H and A is founded on the consistency of the latter. The fact that the GDA is no Krylov method in case of a cycle-dependent shifting strategy is of no relevance for ϕ, but offers the opportunity for an improvement of the global convergence properties.

9.2.15 Residual minimization

Since Section 9.2.5 we have considered possibilities to improve the *acquisition* (→ flexible preconditioning by C) and *maintenance* (→ complete storage) of information during iterative procedures for finding extreme eigenpairs of large, non-factorizable, non-invertible Hermitian operators H. Next we consider a possibility to improve the *use* of information, as well, based on the existence of the alternative "RR or RM" as mentioned in Section 9.2.4.

The RR solves in each cycle S in step (2) the eigenproblem of the $S \times S$ Ritz matrix

$$R_{\mathrm{RR}} = < \mathbf{X}^{(S)} \,|\, H\, \mathbf{X}^{(S)} >$$

to find the vector $\mathbf{v}^{(S)} \in \mathcal{X}$ minimizing the Rayleigh quotient, where $\mathcal{X} = \mathrm{span}\,(\mathbf{X}^{(S)})$. In contrast, the RM determines $\mathbf{v}^{(S)}$ as to minimize the residual norm by solving the eigenproblem of

$$\begin{aligned} R_{\mathrm{RM}} &= < (H - \sigma I)\, \mathbf{X}^{(S)} \,|\, (H - \sigma I)\, \mathbf{X}^{(S)} > && (9.37) \\ &= < \mathbf{X}^{(S)} \,|\, (H - \sigma I)^2\, \mathbf{X}^{(S)} > \qquad . && (9.38) \end{aligned}$$

Thus the RM applied to H is equivalent to the RR applied to the operator

$$K = (H - \sigma I)^2 \qquad , \tag{9.39}$$

meaning that the RM can be introduced to the GDA by formally replacing H with K in step (2). The shifting strategy in step (5b) may nevertheless choose σ as the currently best approximation for λ_1 , regardless of where this approximation is actually taken from. Even for the choice $\sigma = \alpha$, i. e. the current Ritz value (the smallest eigenvalue of R_{RR}), the residuum $\mathbf{r}$ in step (4) obtained subsequent to the RM has a smaller norm than by the RR. Hence, according to Eq. (9.4), the RM effects a decrease of κ and thus an increase of ϕ. This acceleration is the basis of the "direct inversion in the iterative subspace" (DIIS) method [278, 313, 314]. Since K and A are consistent for $\sigma \to \lambda_1$ the RM will not obstruct the assumed validity of Eq. (9.36) either. In particular, by use of Eq. (9.37) for updating R_{RM} the complexity of a cycle is still kept to be $z = 1$.

It has been found, however, that the global convergence towards $\mathbf{u}_1$ of an algorithm entirely replacing the RR with the RM, e. g. the GDA in step (2), may be unreliable [306, 314, 315]. If σ in every cycle is strictly chosen to be the current

Ritz value α the procedure may terminate at a higher eigenvalue than λ_1. We found a different shifting strategy to be advantageous which chooses σ distinctly below the Ritz value α or even pins σ at a constant value $\leq \lambda_1$, in order to achieve global convergence towards $\mathbf{u}_1$. Another possibility to find $\mathbf{u}_1$ is to *combine* the RR and RM during the calculation of $\mathbf{v}^{(S)}$ as suggested in Ref. [306]. Formally, the latter variant employs both H and K in step (2) of the GDA.

9.3 Improved algorithm for interior eigenpairs

All of the methods considered so far find *extreme* eigenpairs only. The single-vector variants yield either $\mathbf{u}_1$ or $\mathbf{u}_N$, while simultaneous variants or variants using deflation techniques [274, 277, 294] either find $\mathbf{u}_1 \ldots \mathbf{u}_M$ or $\mathbf{u}_{N-M+1} \ldots \mathbf{u}_N$, depending on the spectral shift adjustment (in the PM, CI, and CG) or the selection of Ritz values (in the LA and GDA).

This property is entirely based on the particular *selection* of the *wanted* information during the steps labeled by "use of information". In the PM, CI, and CG no selection occurs at all so that, via the gain mechanism explained in Section 9.2.2, the dominance of the extreme eigenvalues determines the *limes* vector $\mathbf{x}^{(\infty)}$ of the respective iteration completely. In contrast, in the LA and all its preconditioned derivates (like the GDA) the selection is made by the RR or RM by *intentionally* extracting the smallest (or largest) Ritz values (from H or K, respectively) in each cycle. The intention in the frame of this work is, however, to extract interior eigenpairs from H, specified by an interval for the eigenvalues only.

The extraction of an approximant to a particular interior eigenvector from the subspace $\mathcal{X}$ by means of the RR is possible if (i) a zeroth-order approximation of that eigenvector is already known [305] or if (ii) $\mathcal{X}$ contains good approximations to all eigenvectors belonging to eigenvalues between the wanted one and the nearest spectral bound of H [306]. In the latter case the performance is based on implicit deflation. If neither (i) nor (ii) is the case the extraction will fail, as a matter of experience, due to the inability to select *the proper Ritz vector* for improvement. To the author's knowledge, it does not work to select a Ritz pair from $R^{(S)}$ in step (2) according to the eigenvalue α only: In practice, such procedures do not converge at all. On the other hand, the existence of possibility (i) suggests that this failure is, in fact, only due to the bad selection criterion: In Ref. [305] nothing but the criterion, which Ritz pair to take in step (2) of the DA, has been changed as compared to the implementation of the GDA described above.

The two criteria applying in cases (i) and (ii) are successful since each of them defines a desired property of the eigen*vector* rather than the eigen*value*. Hence, it is at hand that the selection criterion

$$||\mathbf{r}||_2 \rightarrow \text{min.} \qquad \text{for} \qquad \sigma = E_{\text{search}}$$

in step (4) ensures local convergence towards the wanted, interior eigenpair as well while the wanted eigenvector can be specified via its corresponding eigenvalue by E_{search} [245], as needed for this work. However, as discussed in Section 9.2.15 a total replacement of the RR with the RM in step (2) of the GDA may obstruct the global convergence of the procedure, either by terminating the iteration at an unwanted eigenvector, or by causing a stagnation of the GDA within an endless loop far away from any reasonable approximation of the desired eigenspace. This inclination to failure in connection with the RM is a particular property of the GDA, in connection with different – however, less efficient – basic algorithms the RM does not necessarily cause problems [245, 278].

Basic inventions

To, nevertheless, benefit from both the RM and the GDA, while avoiding the discussed failure sources, *a combination of RM and RR is proposed in this work*:

- In step (2) of the GDA, see p. 147, we replace the RR with the RM, using Eq. (9.37) with fixed $\sigma = E_{\text{search}}$.

  > In contrast, in Ref. [306] $\sigma = \rho_H[\mathbf{v}^{(S)}]$ is used. This dependence introduces the possibility for the GDA to converge to an unwanted eigenvector and necessitates the invention of an appropriate rejection procedure. The present approach to fix σ to the search energy in step (2) securely prevents the algorithm from focussing on undesired eigenvectors.

- In steps (5b) to (9) of the GDA, see p. 147, we use *two* correction vectors:
  (i) The conventional pseudoresidual $\mathbf{y}$ with respect to H and
  (ii) a second pseudoresidual with respect to $(H - E_{\text{search}} I)^2$:

$$\mathbf{y} = C_1^{-1} [\, H - (E_{\text{search}} + \sigma_1)\, I \,]\, \mathbf{v}^{(S)} \qquad \text{and} \tag{9.40}$$

$$\mathbf{z} = C_2^{-1} [\, (H - E_{\text{search}} I)^2 - \sigma_2\, I \,]\, \mathbf{v}^{(S)} \qquad . \tag{9.41}$$

While $\mathbf{y}$ will be calculated and processed in each cycle S, $\mathbf{z}$ will only be used in every ω-th cycle, i. e., if $S \bmod \omega = 0$.

In special cases of H, and in selected cycles, even a third correction vector can be employed. This is particularly advantageous for handling the degeneracy of energy levels due to time reversal symmetry, where an approximant for the second basis vector of an eigenspace can be calculated from a present eigenvector approximant by relation Eq. (4.22).

> Unlike this approach, in Ref. [306] only $\mathbf{y}$ is used, formally corresponding to $\omega = \infty$. *Vice versa*, the folded spectrum method (FSM) [246, 271] uses only $\mathbf{z}$, but never $\mathbf{y}$. Also the RMM-DIIS method [278] uses strictly one correction vector per cycle (which additionally is constructed in a different way).

A complete waive of $\mathbf{z}$ [40, 306] is inacceptable as it enables a stagnation of the GDA within an endless loop caused by numerical linear dependence of $\mathbf{y}$ on $\mathbf{X}$ (in such a case, $\mathbf{b} = \mathbf{0}$ in step (6) prevents further changes of $\mathbf{X}$ and R_{RM}). On the other hand, waiving $\mathbf{y}$ is not known to obstruct the convergence but makes the algorithm converge slower than possible by additionally using $\mathbf{y}$.

The concrete value of ω is adjusted empirically to yield the maximal convergence rate ϕ: While the inclusion of $\mathbf{z}$ to the trial vector basis $\mathbf{X}$ certainly improves the information on the wanted eigenvector, its calculation and the subsequent processing in step (9) enhance the complexity of the respective (ω-th) cycle from $z = 1$ to $z = 3$. Thus the *average* convergence rate is

$$\phi(\omega) \;=\; -\frac{\omega}{\omega + 2} \ln \kappa(\omega) \qquad .$$

The increase of $\kappa(\omega)$ for $\omega > 1$ is partly compensated by the decreasing average cycle complexity, $\overline{z} = (\omega + 2)/\omega$, leading to the existence of an optimal ω value.

Advantage over prevailing methods

The convergence rate of the FSM is $\phi_{\mathrm{FSM}} = -(\ln \kappa_{\mathrm{FSM}})/2$. At least for $\omega = 1$, $\kappa(\omega) \leq \kappa_{\mathrm{FSM}}$ since the inclusion of $\mathbf{y}$ to $\mathbf{X}$ cannot deteriorate the quality of $\mathrm{span}(\mathbf{X})$ as an approximant for $\mathrm{span}(\mathbf{u}_j)$, but usually improves it. As a matter of experience the improved GDA (IGDA) proposed here has a higher convergence rate $\phi(\omega) \geq \phi(1) \geq \phi_{\mathrm{FSM}}$ than the FSM already for $\omega = 1$, inspite of the then 50% larger $\overline{z}$ value. According to practical observation during this work, $\omega = 4$ yields an even faster convergence, as demonstrated below by bench-marking.

The advantage of the IGDA over the FSM may additionally be visualized by the following considerations. The original FSM [246] uses a PCG in the Polak-Ribiere version to calculate the lowest (i. e., extreme) eigenvalues of the spectrally transformed Hamiltonian

$$K = (H - E_{\text{search}} I)^2 \quad . \tag{9.42}$$

Thereby *one* preconditioner can be chosen which corresponds to C_2 in Eq. (9.41) but, in contrast to C_2, needs to be positive definite according to Eq. (9.33). Apart from overcoming this restriction, the IGDA has a second, independent choice C_1 in Eq. (9.40) at hand, thus being more flexible for preconditioning than the FSM.

For the sake of simplicity, and in order to maintain the generality of the considerations, we will neglect preconditioning in the benchmarks and compare only the basic algorithms to each other. Thus, the FSM becomes a CG equivalent[4] to the Arnoldi procedure (see Section 9.2.5) applied to K from Eq. (9.42) instead of H, still having a cycle complexity $z = 2$. In the IGDA, the waive for preconditioning yields $C_1 = C_2 = I$ in Eq. (9.40) and Eq. (9.41), so that

$$\begin{aligned} \mathbf{y} &= \mathbf{w} - (E_{\text{search}} + \sigma_1)\mathbf{v}^{(S)} \qquad \text{and} \\ \mathbf{z} &= H\mathbf{w} - 2E_{\text{search}}\mathbf{w} + ({E_{\text{search}}}^2 - \sigma_2)\mathbf{v}^{(S)} \qquad . \end{aligned}$$

When $\mathbf{y}$ is submitted to the steps (6) to (9) on p. 148 the term containing $\mathbf{v}^{(S)}$ will be deleted by the orthogonalization procedure as it is linearly dependent on $\mathbf{X}$. Hence, $\mathbf{y} = \mathbf{w} = H\mathbf{v}^{(S)}$ is an equivalent choice for the first correction vector. If, thereafter, $\mathbf{z}$ is processed in the same way, the two scalar multiples of $\mathbf{w}$ and $\mathbf{v}^{(S)}$ will vanish accordingly (since $\mathbf{w}$ has already become a member of $\mathbf{X}$). Hence, $\mathbf{z} = H\mathbf{w} = H^2\mathbf{v}^{(S)}$ is an equivalent choice for the second correction vector. Consequently, a shifting strategy for determining σ_1 in Eq. (9.40) and σ_2 in Eq. (9.41) is obsolete in absence of preconditioners C_1 and C_2.

Without preconditioning the IGDA becomes a Krylov method constructing a Krylov space of H while the unpreconditioned FSM constructs a Krylov space of H^2. Hence, the FSM effectively misses each second iterate of the IGDA while, for $\omega \geq 2$, it has at least the same cycle complexity.

The effect of this deficiency can be demonstrated as follows. As an extreme, but simple example we consider the calculation of $\mathbf{u}_1$ (which could be accomplished by the conventional CG or GDA, of course). Thus we may choose $\sigma = E_{\text{search}} = \lambda_1$ in Eq. (9.37), obtaining the spectral gap of K in Eq. (9.42) as $\gamma_K = {\gamma_H}^2$. The convergence rate of the unpreconditioned FSM is then given by Eq. (9.10):

$$\phi_{FSM} \;\geq\; \sqrt{\gamma_K} \;=\; \gamma_H \qquad . \tag{9.43}$$

---

[4] The equivalence of the CG and the Arnoldi procedure is conditional on the use of exact arithmetic in the CG. For our bench-marking we assume this and emulate the CG by an "unpreconditioned GDA" with large truncation intervals (all 200 cycles).

In the considered example, the performance of the FSM is comparable to that of the PM, and so is that of the LA when applied to K [283, 316]. In contrast, the IGDA performs at least as the Arnoldi procedure since it (i) uses the same correction vector as the latter (plus one additional correction vector in one out of ω cycles), and (ii) is convergence-accelerated by the RM Eq. (9.37) using a fixed shift σ. Hence, Eq. (9.10) yields

$$\phi_{\mathrm{IGDA}} \;\geq\; \frac{2\omega}{\omega+2}\sqrt{\gamma_H} \;\geq\; \sqrt{\gamma_H} \qquad \text{for} \qquad \omega \geq 2 \quad . \tag{9.44}$$

The "$>$" relations in Eq. (9.43) and Eq. (9.44) stem from the incorporation of $\mathbf{x}^{(1)}$ into the construction of $V_s(H)$ (see Section 9.2.8), which could not be further quantified. For the purpose of a comparison we might, however, ignore this and replace "$\geq$" with "$\approx$". Thus we conclude that the advantage of the IGDA over the FSM is the larger, the closer to the spectral bounds of H the wanted eigenvalues λ_j are located, i. e. the smaller $\min(\lambda_j - \lambda_1, \lambda_N - \lambda_j)$ is. The relatively slow convergence of the FSM was already observed by its inventors [246].

Implementation

To calculate several interior eigenvectors from H, e. g. the lowest M CB states in a QD, basically a single-vector version of the IGDA is used in connection with implicit deflation [274, 277, 294]. Thereby, for all wanted states the same search energy E_{search} is specified which, in the example, is placed somewhat below the expected CB groundstate energy, but well above the VB groundstate energy. Thus the desired lowest CB states are properly identified (by their energies) while the topmost VB states have energies too far apart from E_{search}.

Then all M wanted (CB) states are iterated until convergence, one after another, beginning with the CB groundstate which is energetically closest to E_{search}. To prevent the m-th iteration, $m = 2, \ldots, M$, from converging to the CB groundstate as well, every (correction) vector to be appended to the trial subspace basis $\mathbf{X}$ is, prior to the appending, orthogonalized against all $m-1$ already calculated (CB) states (implicit deflation). Thus span$(\mathbf{X})$ during the m-th iteration does never contain the $m-1$ already known states, and the iteration will therefore converge to the first *unknown* state having an energy larger than E_{search}. It is advantageous to demand a larger accuracy than needed from the first calculated states, in order to ensure that the finally required accuracy can be obtained, though using deflation, for the highest wanted state, as well.

This concept works, without restrictions, as well if energy levels are degenerate, as in the case of time reversal symmetry. Then for each eigenspace an orthonormal basis is obtained, whereby the geometric multiplicity of each energy level can be *determined* from the calculation. Hence accidental degeneracies of confined levels are always properly recognized.

Particularly in the presence of time reversal symmetry the calculation of each second state is practically costless, as an outstanding approximate for the second basis vector of the respective eigenspace can be obtained from the already known, first basis vector by using Eq. (4.22).

In practice, the true implementation used in this work (and in the benchmarks) enables phases of simultaneous iteration where, for instance, the two basis vectors for the CB groundstate eigenspace still are being iterated for higher final accuracy, while the iteration for the first excited CB state has already begun, being in the initial, global phase of convergence. Such phases are visible in the iteration histories shown in Fig. 9.3 and Fig. 9.2. However, the description of details concerning the optional simultaneous process would make the implementation outline given below excessively complicated without being of fundamental interest. Therefore, only the single-vector variant of the IGDA for the calculation of one interior eigenpair is presented here, neglecting even the recommended exhaustion of time reversal symmetry, if applicable.

0.) If $S = 0$ then

   (a) an N-dimensional initial guess vector $\mathbf{x}^{(1)}$ is (i) random-generated, or (ii) calculated from another eigenvector using, e. g., Eq. (4.22), or (iii) taken from elsewhere, and stored,

   (b) its image $\mathbf{h}^{(1)} = (H - E_{\text{search}} I)\mathbf{x}^{(1)}$ is calculated and stored,

   (c) the scalar product $\mathbf{h}^{(1)\dagger}\mathbf{h}^{(1)}$ is calculated and stored as the (1,1) matrix element of the 1×1 RM matrix $R^{(1)}$,

   (d) $S = 1$.

   (e) $\Omega = 0$.

   Otherwise, the two sequences of the previous trial vectors $\mathbf{x}^{(s)}$ and their images $\mathbf{h}^{(s)} = (H - E_{\text{search}} I)\mathbf{x}^{(s)}$ $(s = 1, \dots, S)$ are prevailing and form the columns of the two $N \times S$ matrices $\mathbf{X}^{(S)}$ and $\mathbf{H}^{(S)}$, respectively.

   The $S \times S$ RM matrix $R^{(S)} = \mathbf{H}^{(S)\dagger}\mathbf{H}^{(S)}$ is also prevailing.

1.) If there exists a prevailing RM vector from the previous cycle it is stored as a back-up potentially needed during a basis truncation in the following step.

2.) RM: The RM vector $\mathbf{v}^{(S)} = \sum_{s=1}^{S} a_s \mathbf{x}^{(s)}$ is calculated from the S-dimensional eigenvector a belonging to the lowest eigenvalue α of $R^{(S)}$.

If storing all forthcoming correction vectors and their images would exceed the storage capacity then

(a) all $\mathbf{x}^{(s)}$ and $\mathbf{h}^{(s)}$ are disposed ($s = 1, \ldots, S$),

(b) the back-up RM vector is orthonormalized against $\mathbf{v}^{(S)}$,

(c) both vectors are stored as $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, respectively,

(d) their images $\mathbf{h}^{(1)}$ and $\mathbf{h}^{(2)}$ are calculated explicitly and stored,

(e) the 2×2 RM matrix $R^{(2)}$ is calculated,

(f) $S = 2$.

(g) $\Omega = 0$.

Otherwise the RM vector back-up is disposed and Ω is incremented by one.

3.) The image $\mathbf{w} = \sum_{s=1}^{S} a_s \mathbf{h}^{(s)}$ of the RM vector $\mathbf{v}^{(S)}$ and the corresponding Rayleigh quotient $\rho = \mathbf{w}^\dagger \mathbf{v}$ are calculated.

4.) The residual $\mathbf{r} = \mathbf{w} - \rho \mathbf{v}^{(S)}$ is calculated. If $||\mathbf{r}||_2 <$ tol the IGDA terminates.

5.) σ_1 and σ_2 are chosen according to a shifting strategy, for instance, $\sigma_1 = \rho$ and $\sigma_2 = {\sigma_1}^2$.

6.) The first correction vector

$$\mathbf{y} \; = \; {C_1}^{-1} \, [\, \mathbf{w} - \sigma_1 \mathbf{v}^{(S)} \,]$$

is calculated, orthonormalized against $\mathbf{X}^{(S)}$, and then appended to the latter as a new column $\mathbf{x}^{(S+1)}$ yielding $\mathbf{X}^{(S+1)}$. The image of $\mathbf{x}^{(S+1)}$ is calculated and appended as a new column $\mathbf{h}^{(S+1)}$ to $\mathbf{H}^{(S)}$ yielding $\mathbf{H}^{(S+1)}$. The RM matrix $R^{(S)}$ is updated by adding a new column $c = {\mathbf{H}^{(S+1)}}^\dagger \mathbf{h}^{(S+1)}$ as well as a new row $c^\dagger$, yielding $R^{(S+1)}$, and S is incremented by one.

7.) If $\Omega \bmod \omega = 0$ then

the second correction vector

$$\mathbf{z} \; = \; {C_2}^{-1} \, [\, (H - E_{\text{search}} I)\mathbf{w} - \sigma_2 \mathbf{v}^{(S-1)} \,]$$

is calculated[5] and further processed like $\mathbf{y}$ in step (6).

8.) The next cycle starts with step 1.

---

[5] Note that the same RM vector $\mathbf{v}$ as in step (6) is used here since the cycle counter S has already been incremented.

The termination criterion in step (4) in principle permits the IGDA to terminate at a higher eigenpair than the wanted, as well, since strictly the Rayleigh quotient ρ is used to calculate the residual norm. However, as a matter of experience the global convergence to the eigenpair closest to E_{search} is ensured by the incorporation of $\mathbf{z}$ in step (7).

The suggestion to choose $\sigma_2 = \rho^2$ in step (5) is based on the fact that this is generally is a better approximation to the smallest eigenvalue of $(H - E_{\text{search}} I)^2$ than α: Note that $\alpha = \rho^2 + ||\mathbf{r}||_2{}^2$ (which, in practice, is the most sensitive check whether H is truly Hermitian).

9.3.1 Benchmarks

The efficiency of the IGDA was explained in theory, giving various arguments that the IGDA has a higher convergence rate than prevailing methods used in the field for the calculation of interior eigenpairs (apart from shift-and-invert algorithms which are impractical in the given case). Finally, benchmarks are presented to contrast the IGDA with the FSM which is considered the most efficient alternative method. In order not to loose generality by considering certain preconditioners, the basic, unpreconditioned algorithms are going to be compared here.

The unpreconditioned IGDA, defined by $C_1 = I$ and $\sigma_1 = 0$ in step (6) and $C_2 = I$ and $\sigma_2 = 0$ in step (7), becomes equivalent to the unpreconditioned FSM (in exact arithmetic) by (i) choosing $\omega = 1$ and (ii) omitting step (6), as explained in Section 9.3. Thus, without the implementation of a true CG procedure, an indirect comparison of both methods becomes possible which, at least, is reliable with respect to the convergence rate. Since both methods take into account the spectral properties of H in approximately the same manner, the comparison may even be considered to characterize the difference between the IGDA and the FSM in general.

As a practical example for the comparison of the (unpreconditioned variants of) IGDA and FSM the electronic eigenstates of the QD from Chapter 3 as shown in Fig. 3.1(c) are calculated. Thereby two different FDM grid spacings d are compared in order also to demonstrate the impact of the related spectral structure of H on the convergence rate: The "coarse" grid uses $d = a$ (as throughout Part II) with a being the lattice constant, while the "fine" grid employs the original resolution of the calculations in Chapter 3 (and in set 1 of Table 4.3), $d = a/2$. The differences between corresponding eigenpairs calculated using both grid resolutions are small, as discussed in Section 4.3.2. In all benchmark calculations the bases $\mathbf{X}$ and $\mathbf{H}$ were truncated at $S = 200$ in step (2) in order to keep the influence of truncation on the convergence properties minimal.

On the coarse grid ($d = a$) the QD and the surrounding barrier are represented by $36 \times 36 \times 30$ voxels yielding a real Hamiltonian rank of $N = 622080$ (38880 grid points to all of which eight complex variables for the eight bands are assigned). The calculation of the lowest (highest) four pairwise degenerate CB (VB) levels, like those shown in Fig. 3.1(c), requires two runs with different search energies E_{search} yielding eight eigenvectors, each. The search energies for both runs, given on the same scales as in Fig. 3.1, are $E_{\text{search}} = -460\,\text{meV}$ for the CB states and $E_{\text{search}} = 258\,\text{meV}$ for the VB states. The iteration histories of the two runs are shown in Fig. 9.2 displaying the residual norm $||\mathbf{r}||_2$ in step (4) of the IGDA (or the modified IGDA simulating the FSM, respectively) over the number of performed $H\mathbf{x}$ multiplications. From such plots the convergence rate is easily determined directly.

The accuracy goal "tol" for $||\mathbf{r}||$ in step (4) is $0.1\,\text{meV}$ and indicated in the plots by a solid line. With this residual tolerance for the eigen*vectors* the eigen*values* are, in practice, converged to full (single) machine precision, corresponding to $\sim 0.1\,\mu\text{eV}$. However, to calculate $M = 8$ eigenvectors using the same search energy in connection with deflation the accuracy goal of the m-th eigenvector ($m = 1, \ldots, M$) is set to $2^{(M-m)/2} \cdot \text{tol}$, as explained above. Therefore all states but the highest (lowest) wanted CB (VB) state are continued to be iterated even for $||\mathbf{r}|| < \text{tol}$. Nevertheless, a new state vector is added for simultaneous processing as soon as tol has been reached. Since, in this example, each second eigenvector approximant can be calculated from a prevailing approximant by means of Eq. (4.22), tol is usually achieved already for the initial guess, hence a third state vector is added immediately. The temporarily simultaneous iteration of thus three vectors manifests itself in a kink of the residual norm curve at $||\mathbf{r}|| \approx \text{tol}$, as three vectors require three times as much $H\mathbf{x}$ multiplications to get improved as compared to a single vector. The convergence rate *of each vector* is, however, hardly influenced by the simultaneous iteration indicating that a block variant offers no advantages concerning the local convergence rate [306].

From comparing the FSM (gray squares) and the IGDA with $\omega = 4$ (black circles) in Fig. 9.2 the superior performance of the latter is well seen. Second, the advantage of the IGDA over the FSM is the larger, the more dense the Hamiltonian spectrum is in the region of interest. The average separation of the CB levels is about five times larger than that of the VB levels, hence the VB states require 2.5 (1.8) times more $H\mathbf{x}$ multiplications than the CB states when using the FSM (IGDA). The difference between the two factors reveals the different scaling behaviours of the convergence rate ϕ with the spectral gap $\gamma_j = (\lambda_{j+1} - \lambda_j)/\Lambda$ in both methods.

Another decrease of γ_j by a factor ~ 4 is obtained when changing from the coarse to the fine grid. Thereby, the explicit dependence of the second order difference quotients of the FDM (see Section 8) on d causes an increase of the spread Λ of H by a factor ~ 4 while the energy level separations $\lambda_{j+1} - \lambda_j$ in the

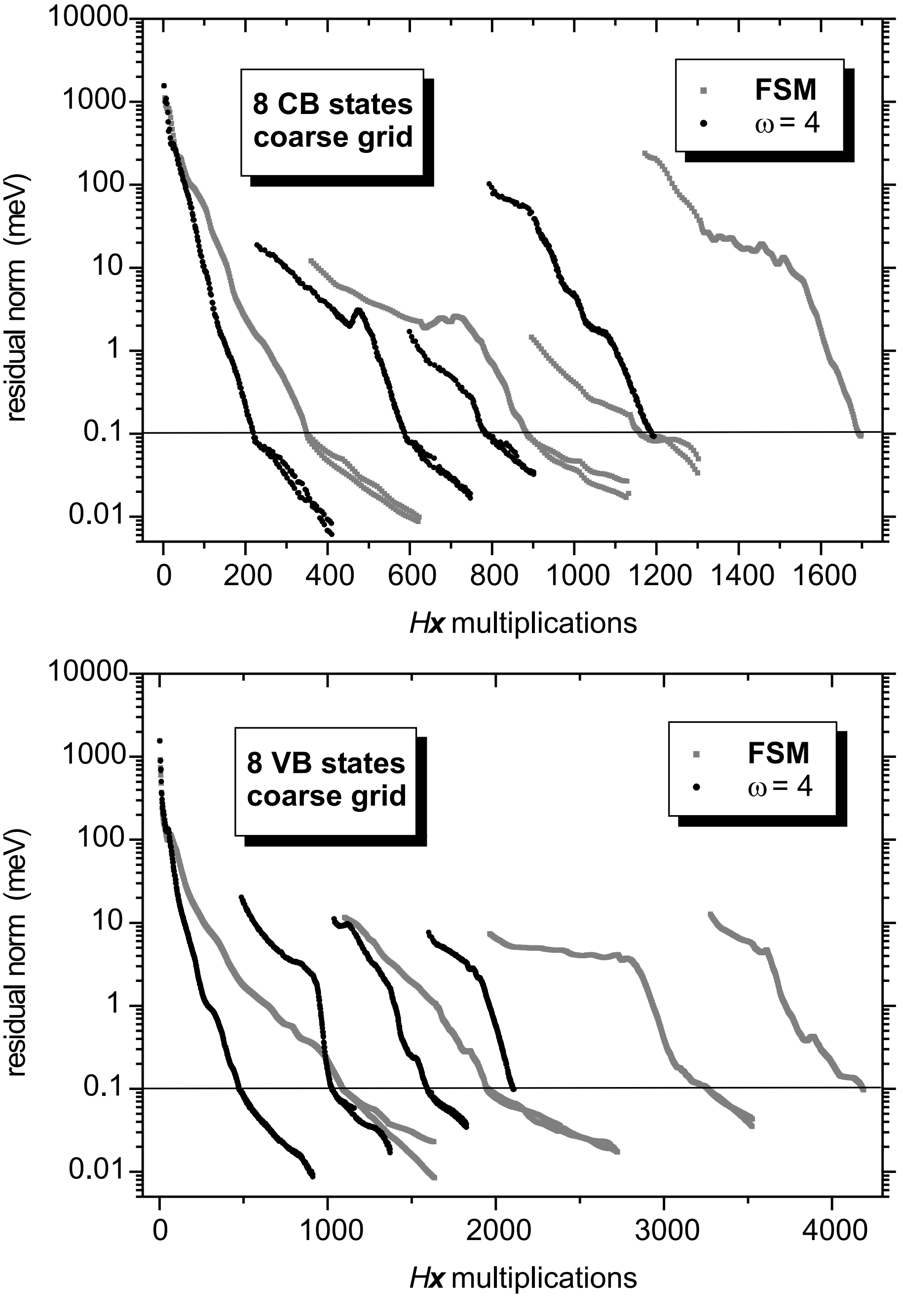


Figure 9.2:

Iteration history of the calculation of eight CB and VB states, each, in the model QD from Chapter 3 using a grid spacing equal to the lattice constant. Black circles show the residual norm of the eigenvector approximations in step (4) of the IGDA (see p. 158) over the number of performed $H\mathbf{x}$ multiplications for $\omega = 4$ (see p. 154). Gray squares refer to the FSM emulation described in the text on p. 159.

QD remain essentially unaltered. The FSM reacts to this with a multiplication count increased by an average factor 2.6 while the IGDA multiplication count increases by a factor 2.1, see Fig. 9.3. On the fine grid ($d = a/2$) the structure is represented by $72 \times 72 \times 60$ voxels yielding a Hamiltonian rank of $N = 4976640$.

The gray circles in Fig. 9.3 show that the IGDA is faster than the FSM even for $\omega = 1$, i. e. the decrease of the convergence factor κ is more significant than the increase of the average cycle complexity $\overline{z}$ (see Section 9.3) when changing from the FSM to the IGDA.

9.3.2 Programming considerations

The bandstructure calculations for QWRs were carried out on DEC Alpha AXP 600 workstations with clockrates between 266 and 433 MHz [40] using DEC Pascal source code written by the author.

The QD eigenstate calculations, in particular the above benchmarks, were performed on a Cray T3E-1200LC136/900LC272 parallel computer[6] (see Fig. 9.4) operated by Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB).

The software for the electronic structure calculation in QD heterostructures was written and developed by the author and consists of three parts. The first part runs on a Compaq XP 1000 workstation (formerly DEC Alpha AXP 600, clock rate 500 MHz, 1.1 GB RAM[7]) and (i) preprocesses any structural information about the QD, (ii) compiles code files defining the bandstructure model to be used (the eight-band **k·p** model in our case), and (iii) generates one file containing a sparse representation the discrete Hamiltonian as well as the value of E_{search}. This program part is encoded in DEC Pascal. The Hamiltonian data file is passed to the T3E parallel computer for eigenvector calculation.

The code for this second part of the software running under the operating system UNICOS/mk 2.0.5.33 is written in Cray Fortran 90 (used for all numerical core routines, like differential operators and vector algebra) and "`lint`-free"[8] ANSI C (used for everything else, like I/O[9], memory management, and the algorithmic administration). Indeed, Fortran yields considerably faster code for the numerics than C in our case, due to the *call by reference* convention for subroutine parameters in connection with the requirement for explicit specification of pointer targets, which in combination enable a more efficient functional-unit

---

[6] Produced by the Silicon Graphics company Cray Research Inc. (CRI).

[7] random access memory

[8] `lint` is a programming accessory (C program checker) on UNIX systems used (i) to test the ANSI standard conformity of C source code and (ii) to discover commonly made mistakes which would not be objected by the compiler as they do not violate language standards.

[9] Input/Output

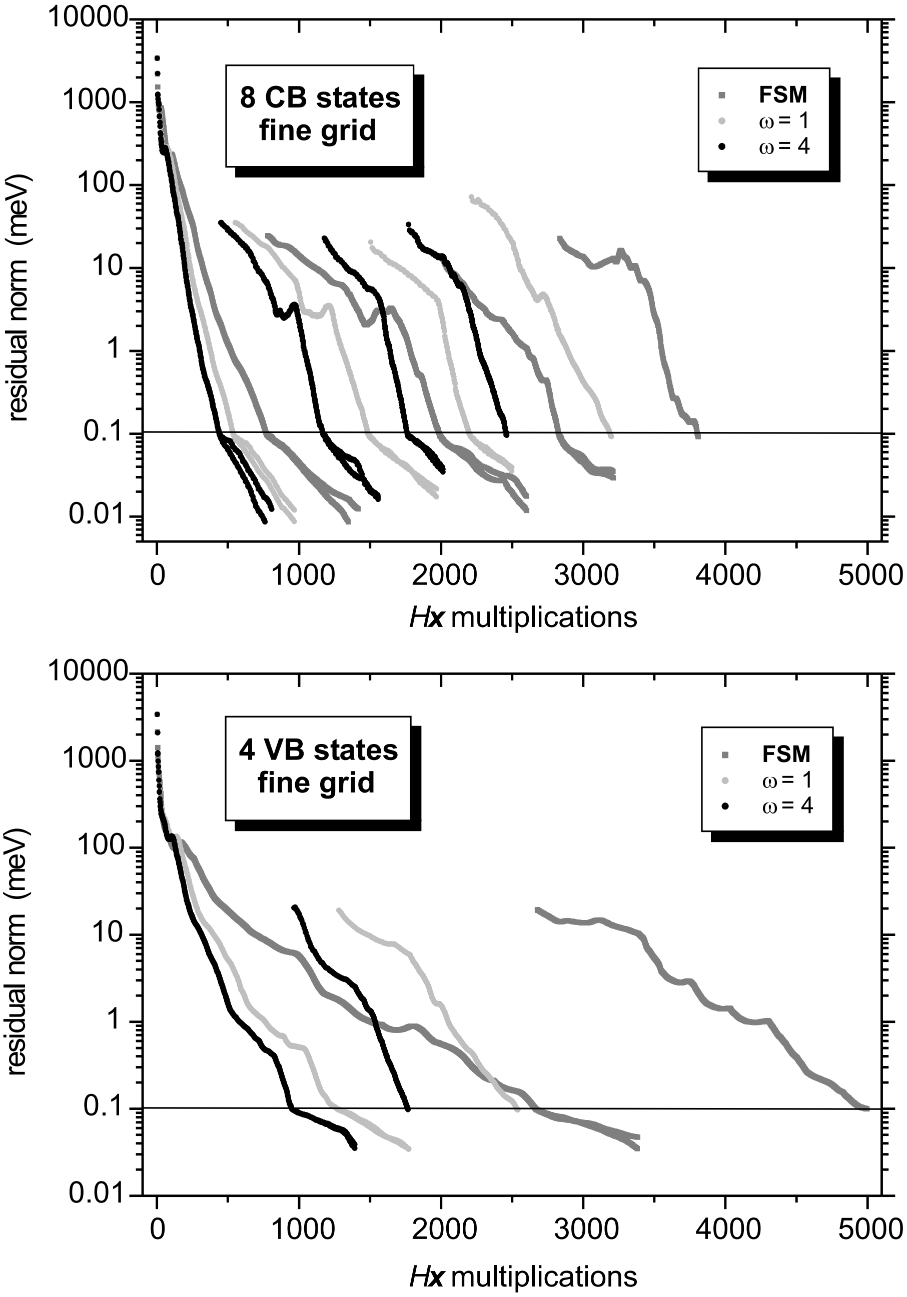


Figure 9.3:

Iteration history of the calculation of eight CB and four VB states in the model QD from Chapter 3 using a grid spacing equal to half the lattice constant. Black (lightgray) circles show the residual norm of the eigenvector approximations in step (4) of the IGDA (see p. 158) over the number of performed $H\mathbf{x}$ multiplications for $\omega = 4$ ($\omega = 1$) (see p. 154). Gray squares refer to the FSM emulation described in the text on p. 159.

Figure 9.4:

© CRI. Scalable parallel computer Cray T3E [317, 318]: Up to 2048 processing elements (nodes), based on the DEC Alpha 21164 microprocessor [254] with direct mapped first-level cache and equipped with a special support circuitry featuring stream detection for accelerated unit-stride access to the local RAM (up to 875 MB/s memory bandwidth), are connected within a three-dimensional torus with an interprocessor network memory bandwidth of up to 350 MB/s (for comparison: current hard-disk drives used in personal computers offer bandwidths of $\sim$ 20 MB/s). Depending on the clock rate (450 MHz or 600 MHz) each node delivers a peak performance of 900 or 1200 millions of 64-bit floating point operations per second (flops), realized if all data are kept in registers. The T3E is a "multiple instruction multiple data" (MIMD) architecture, i. e., each node executes an individual instruction code using local data. The torus operation is coordinated by, e. g. eight, command nodes. The machine "berte" in the ZIB (`berte.zib.de`), used in this work for the electronic structure calculations of QD heterostructures, has 272 nodes with 450 MHz clock rate and 136 nodes with 600 MHz, forming a heterogeneous cluster with 70 GB distributed RAM and an application peak performance of 0.38 Teraflops. In the worldwide supercomputer performance ranking (`www.top500.org`) including also military facilities "berte" kept rank 54 in 06/2000.

optimization by the compiler [254]. The parallelization of the program is based on geometrical domain decomposition within two spatial directions, partitioning the QD heterostructure into congruent rectangular columns each of which being processed by one T3E node. The interprocessor communication is encoded using the Message Passing Interface (MPI) language and needed for the updating of common boundary points as well as for global operations, like scalar products and temporal synchronization. The number of nodes to be used is chosen at run time and depends on the desired speed, the overall RAM size required, and the grid size: The partition columns should not have a too large base area in order to enable a sufficient array blocking [254] to support optimal cache performance.

The memory management of this program is optimized with respect to the first-level data cache use in order to enable an as large as possible data load rate from the RAM. This is important since the RAM access imposes the principal limitation to the overall performance of the program: the CPU[10] of a node usually spends more time waiting for data to be fetched from or put to the RAM, than for doing the relatively few arithmetic operations on them. The eigenvector calculation employs, to almost 100%, typical vector operations (one $H\mathbf{x}$ multiplication may consist of, e. g., 64 multiplications with diagonal matrices, 72 first order and 72 second order differentiations applied to vectors having as much components as the grid has mesh points, as well as a corresponding number of vector additions) while the T3E is not truly specialized to vector calculations. The performance of the given code on the given machine is limited by the achieved rate of cacheable stores.

To widen this bottle-neck, the special stream detection hardware of the T3E nodes efficiently supports strictly sequential (unit-stride) RAM accesses by accelerated feeding of the first-level cache using a read-ahead hardware feature [318]. The first-level cache is backed by a three-way set-associative [254] second-level cache before requests go off-chip to the physical RAM. Making possibly ample use of this circuitry layout requires to strictly avoid cache conflicts and, via the direct mapping [254] of the RAM to the first-level cache, imposes conditions on the address alignment of coherent data blocks being processed simultaneously. These requirements aim at initiating a possibly large number of data *streams* [318] and lead to the set-up of different vector ”classes” being aligned to different first-level cache lines. Since the vector size (being common to all classes) shall be flexible it is unknown at compile-time, so that the block alignment (”padding”) is explicitly organized by the program at run-time. As result of such a memory management the $H\mathbf{x}$ multiplications perform with *average*[11] *sustained*[12] flops rates between ~ 90 and ~ 170 Mflops per node, depending on the number of used nodes (typically between 8 and 100) and their clock rates. The second order difference quotient operators yield up to ~ 280 Mflops per node which means that, in this case, a linear algebra application reaches unusual 23% of the theoretically possible peak performance.

The third and last part of the software, running again on the DEC/Compaq workstation and being written in DEC Pascal, converts the eigenvectors retrieved from the T3E computer to another format making them appropriate for visualization and passing to further software (for the calculation of momentum matrix elements and few-particle state properties) and links them to the input data from the first program, thus generating a complete project description of the QD heterostructure on the single-particle level.

---

[10] central processing unit

[11] The average is taken over all different vector operations involved in one $H\mathbf{x}$ multiplication.

[12] The rate represents an average throughout the execution time for a complete $H\mathbf{x}$ multiplication.

9.4 Resumé

A comparative efficiency analysis of iterative eigensolution methods used in current bandstructure calculations of quantum confined nanostructures has been given. It covers, as basic algorithms, the power method, Chebyshev iteration, conjugate-gradients method, Lanczos algorithm, and the generalized Davidson algorithm, including several aspects of preconditioning. Among others, the molecular dynamics methods of first and second order were shown to be preconditioned variants of the power method and the Chebyshev iteration, respectively. *En passant*, an uncommon approach to overcome the poor local convergence of the conventional second order molecular dynamics method as an eigensolver was found.

A novel, improved variant of the generalized Davidson algorithm, specially tailored for the calculation of interior eigenpairs – as necessary in realistic bandstructure calculations – has been developed, analyzed theoretically, and benchmarked. The algorithm labeled IGDA (improved generalized Davidson algorithm) relies strongly on residual minimization. It serves the same purpose as the folded spectrum method [246] and the spectral transformation Lanczos procedure [316] but was shown to be more efficient than both. The IGDA was hence used throughout this work. Its implementation has been described in detail, taking care of selected programming concerns.

The IGDA can readily be applied to other Hamiltonians describing single particles in realistic nanostructures as well, e. g. in the frame of tight-binding or pseudopotential calculations. The computational progress established by this work is not restricted to the $\mathbf{k}\cdot\mathbf{p}$ envelope function approach.

9.5 Summary and Outlook

In this work, eight-band **k·p** theory including strain and piezoelectricity has been applied to the calculation of the electronic and optical properties of strained quantum dots and quantum wires of realistic shape and chemical composition. Based on the electronic structure, absorption spectra and the properties of few-particle states were calculated and compared to available experiments, demonstrating outstanding agreement between calculated and measured quantities. The most important, overall achievement by this work is that a tractable, computational framework was established whose accuracy has been validated by comparison with independent second-principles calculations. Since no adjustable parameters are employed (except from insufficiently accurate bulk material properties) the theoretical framework outlined here possesses a generic predictive power.

The success of the presented calculations is based on the use of a theoretical model which, first of all, allows taking into account a realistic real structure for the quantum dot or wire, concerning both the geometrical shape and the chemical composition of the entire heterostructure. As specific nanostructures, self-organized InAs quantum dots in GaAs, grown by the Stranski-Krastanow mode, and V-groove quantum wires of InGaAs fabricated by growth on prepatterned substrates of AlGaAs or InP have been modeled.

To model a quantum wire, its cross section shape as known from transmission electron microscopy, as well as its nominal chemical composition are directly used to calculate the conduction and valence band structures in the transport direction. From these, the one-dimensional joint density of states is calculated which governs the structure of the absorption spectrum. These calculations were the first of their kind and published in 1996.

The irregular shape of the wire, its inhomogeneous strain distribution, and the piezoelectric effect create a confinement situation which markedly differs from what would be expected considering solely the geometrical shape seen in microscopy. The overall structure of the one-dimensional electronic subbands results in a recognizable way from the three above mentioned impact factors, which directly visualizes the influence of the structural properties on the electronic and optical properties. The polarizations of optical absorption in strained and unstrained quantum wires differ significantly. These findings indicate that less advanced theoretical approaches fail to describe the complex interrelation between the real structure and the electronic structure. A comparison with photoluminescence excitation spectroscopy experiments demonstrates the high accuracy of the present calculations.

The Stranski-Krastanow grown InAs quantum dots are in the order of 10 nm large so that the quantization energies of confined charge carriers are much larger than in the wires. The predictive power of the eight-band $\mathbf{k}\cdot\mathbf{p}$ model also for such structures is ensured by a deliberate choice of the material parameters for the bulk semiconductors: Given that the $\mathbf{k}\cdot\mathbf{p}$ calculated bulk bandstructure describes the actual bulk bandstructure well up to sufficiently high energies and wave numbers, also the predicted zero-dimensional states in quantum dots are modeled reliably, as demonstrated for the first time in this work. A direct comparison with recent empirical pseudopotential calculations proves the accuracy of the eight-band $\mathbf{k}\cdot\mathbf{p}$ calculation to be $\sim 10\,\mathrm{meV}$. This renders the eight-band $\mathbf{k}\cdot\mathbf{p}$ model fully valid for the theoretical analysis of expitaxial, strained quantum dots.

The electronic properties of InAs quantum dots were calculated exemplarically assuming a pyramidal dot shape. Most probably, this is an idealization of the true shape. However, structural investigations of self-organized InAs/GaAs quantum dots suggest a variety of possible geometries without decisively favoring one shape. The assumption of an InAs pyramid for the dot shape introduces, though in agreement with transmission electron miscroscopy results, one of the greatest uncertainties to the calculations and limits their comparison to experiments *a priori*.

Based on the pyramid model, the electronic level structure in the dots was calculated, examining also the influence of material parameter uncertainties and of different strain models on the calculated bound states. The confined wave functions of electrons and holes in these quantum dots exhibit a pronounced mixing of the Brillouin zone center bulk bands, preventing their classification as, e. g., heavy holes or light holes. This has consequences on the fine-structure of confined excitons in such dots: Even in absence of magnetic fields the fine-structure of the exciton ground state accommodates no degenerate levels. The predicted exciton level splitting due to exchange effects agrees excellently with single-quantum-dot spectroscopy results.

The calculation of few-particle states in InAs/GaAs quantum dots was accomplished by a configuration interaction treatment, including the long-range exchange effects aside of the direct Coulomb interaction and correlation. Thus, trions and biexcitons were modeled, and it was shown that the binding energies of these states sensitively depend on the quantum dot shape and size. The binding energies of biexcitons can be positive or negative, in agreement with single-quantum-dot spectroscopy results.

Optical absorption spectra due to excitons or intraband transitions were calculated in the dipole approximation. The pronounced dependencies of the spectra on the dot size or the polarization direction are clearly related to the real structure of the quantum dot via the wave-function shapes. This is particularly well seen on the enhancement of the polar exciton-phonon interaction in piezoelectrically charged quantum dots which, for the first time, was predicted in the frame of this work in ample agreement with experiments. Like in the quantum wire case, the calculation directly visualizes the impact of the dot geometry on the optical properties, enabling conclusions being partly transferable to other QD systems.

The present calculations for InAs/GaAs quantum dots also contributed to the first unambiguous proof of the quantum size effect in self-organized, Stranski-Krastanow-grown quantum dots. Spectroscopy on various high-quality quantum dot samples largely confirms the predicted dependence of excited exciton state transition energies on the quantum dot size, indicating at the same time that the major limitation of the theory results from lacking information on the real structure and chemical composition of the dots.

The application of the presented approach to forthcoming systematic investigations of the correlation between the structural and optical properties of low-dimensional semiconductor structures relies on the use of supercomputing resources. As a secondary achievement within this work, the currently fastest existing algorithm for the calculation of interior eigenpairs from very large, non-transposable, non-factorizable, non-invertible, Hermitian matrices has been developed. Benchmarks prove the enhanced performance of the procedure, whose implementation is described in detail, allowing its adaption to other work. As a specific implementation, the algorithm was encoded for a parallel computer of the type Cray T3E.

Acknowledgements

I am grateful to Prof. Dr. Dieter Bimberg who gave me the opportunity to learn, teach, and work in his group during eight years, and who persistently stimulated and supported the progress of my work.

Prof. Dr. Ortwin Hess (Deutsches Zentrum für Luft- und Raumfahrt) and Priv.-Doz. Dr. Axel Hoffmann are cordially acknowlegded for refereing this thesis.

My sincere gratitude is due to Dr. Robert Heitz who is the ideal partner for a cooperation having led to a substantial part of the achievements reported here.

During the largest part of the way leading to this thesis I was accompanied by Dr. Volker Türck who provided any, whatsoever support, for instance: Encouragement, fruitful discussions, and assistence in all computer concerns.

Prof. Dr. Marius Grundmann (now at Universität Leipzig) has developed all the software for strain calculations used in this work and is acknowledged for plenty of important suggestions.

Andrei Schliwa was a valuable help in the modeling of few-particle states.

A large part of the calculations were performed on the Cray T3E of Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) within grant Bvpt13.

Exceptionlessly without delay, waiving any bureaucratic attitude, Dr. Hinnerk Stüben, project consultant at ZIB, supported my work on the Cray T3E. I am very thankful for his help.

Dr. Armin Dadgar is acknowledged for the reliable administration of the group's Windows-NT network.

I thank my parents Achim and Helga Stier for always having encouraged me to follow my own way.

And Christiane Schulz for being there.

Bibliography

[1] D. Bimberg, M. Grundmann, N. N. Ledentsov, Quantum Dot Heterostructures (J. Wiley & Sons, Chichester, 1999).

[2] L. Jacak, P. Hawrylak, A. Wójs, Quantum Dots (Springer, Berlin, 1998).

[3] S. V. Gaponenko, Optical Properties of Semiconductor Nanostructures (Cambridge University Press, Cambridge, USA, 1998).

[4] U. Woggon, Optical Properties of Semiconductor Quantum Dots (Springer Tracts in Modern Physics Vol. 136, Berlin, 1997).

[5] L. Bányai, S. W. Koch, Semiconductor Quantum Dots (World Scientific Series on Atomic, Molecular and Optical Properties Vol. 2, Singapore, 1993).

[6] M. Grundmann, E. Kapon, J. Christen, D. Bimberg, Int. J. of Nonlinear Optical Physics **4**, 99 (World Scientific, Singapore, 1995).

[7] E. Kapon, Semiconductors and Semimetals, **40**, 259 (1994).

[8] E. Kapon, Proc. IEEE, **80**, 398 (1992).

[9] S. Raymond, S. Fafard, P. J. Poole, A. Wojs, P. Hawrylak, S. Charbonneau, D. Leonard, R. Leon, P. M. Petroff, J. L. Merz, Phys. Rev. B **54**, 11548 (1996).

[10] K. Mukai, N. Ohtsuka, H. Shoji, M. Sugawara, Phys. Rev. B **54**, R5243 (1996).

[11] P. W. Fry, I. E. Itskevich, D. J. Mowbray, M. S. Skolnick, J. J. Finley, J. A. Barker, E. P. O'Reilly, L. R. Wilson, I. A. Larkin, P. A. Maksym, M. Hopkinson, M. Al-Khafaji, J. P. R. David, A. G. Gullis, G. Hill, J. C. Clark, Phys. Rev. Lett. **84**, 733 (2000).

[12] R. Heitz, O. Stier, I. Mukhametzanov, A. Madhukar, D. Bimberg, Phys. Rev. B **62**, 11017 (2000).

[13] G. Goldoni, F. Rossi, E. Molinari, A. Fasolino, R. Rinaldi, R. Cingolani, Appl. Phys. Lett. **69**, 2965 (96).

[14] F. Vouilloz, D. Y. Oberli, M.-A. Dupertuis, A. Gustafsson, F. Reinhardt, E. Kapon, Phys. Rev. Lett. **78**, 1580 (1997).

[15] E. Martinet, F. Reinhardt, A. Gustafsson, G. Biasiol, E. Kapon, Appl. Phys. Lett. **72**, 701 (1998).

[16] E. Martinet, M.-A. Dupertuis, F. Reinhardt, G. Biasiol, E. Kapon, O. Stier, M. Grundmann, D. Bimberg, Phys. Rev. B **61**, 4488 (2000).

[17] M. Grundmann, Physica E **5**, 167 (2000).

[18] D. Bimberg, N. Kirstaedter, N. N. Ledentsov, Zh. I. Alferov, P. S. Kop'ev, V. M. Ustinov, IEEE J. Sel. Top. Quantum Electron **3**, 196 (1997).

[19] V. A. Shchukin, D. Bimberg, Rev. Mod. Phys. **71** No. 4, 1125 (1999).

[20] L. R. C. Fonseca, J. L. Jimenez, J. P. Leburton, R. M. Martin, Phys. Rev. B **57**, 4017 (1998).

[21] L. R. C. Fonseca, J. L. Jimenez, J. P. Leburton, Phys. Rev. B **58**, 9955 (1998); Phys. Rev. B **60**, 2127 (1999).

[22] S. Nagaraja, J. P. Leburton, R. M. Martin, Phys. Rev. B **60**, 8759 (1999).

[23] J. A. Barker, E. P. O'Reilly, Phys. Rev. B **61**, 13840 (2000).

[24] M. A. Cusack, P. R. Briddon, M. Jaros, Phys. Rev. B **54**, R2300 (1996).

[25] H. Jiang, J. Singh, Appl. Phys. Lett. **71**, 3239 (1997).

[26] H. Jiang, J. Singh, Phys. Rev. B **56**, 4696 (1997).

[27] H. Jiang, J. Singh, Physica E **2**, 614 and 720 (1998).

[28] H. Jiang, J. Singh, IEEE J. Quant. Elec. **34**, 1188 (1998).

[29] C. Pryor, Phys. Rev. B **57**, 7190 (1998).

[30] C. Pryor, Phys. Rev. Lett. **80**, 3579 (1998).

[31] O. Stier, M. Grundmann, D. Bimberg, Phys. Rev. B **59**, 5688 (1999).

[32] C. Pryor, Phys. Rev. B **60**, 2869 (1999).

[33] D. Gershoni, C. Henry, G. Baraff, IEEE J. Quant. Elec. **29**, 2433 (1993).

[34] Y. Zhang, Phys. Rev. B **49**, 14352 (1994).

[35] T. Saito, J. N. Schulman, Y. Arakawa, Phys. Rev. B **57**, 13016 (1998).

[36] A. Zunger, MRS Bull. **23**, 35 (1998).

[37] J. Kim, L.-W. Wang, A. Zunger, Phys. Rev. B **57**, R9408 (1998).

[38] L.-W. Wang, J. Kim, A. Zunger, Phys. Rev. B **59**, 5678 (1999).

[39] A. J. Williamson, A. Zunger, Phys. Rev. B **59**, 15819 (1999).

[40] O. Stier, D. Bimberg, Phys. Rev. B **55**, 7726 (1997).

[41] M. L. Cohen, J. R. Chelikowsky, Electronic Structure and Optical Properties of Semiconductors, Solid State Sciences **75** (Springer, Berlin, 1988).

[42] P. Y. Yu, M. Cardona, Fundamentals of Semiconductors (Springer, 1996).

[43] L.-W. Wang, A. Zunger, Phys. Rev. B **51**, 17398 (1995).

[44] Quantum Theory of Real Materials, ed. by J. R. Chelikowski and S. G. Louie (Kluwer Academic, Dordrecht, The Netherlands, 1996).

[45] Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology, New Series (Springer, Berlin, 1982), Vol. III/17a.

[46] J. R. Chelikowski, M. L. Cohen, Phys. Rev. B **14**, 556 (1976).

[47] L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, D. A. Shirley, Phys. Rev. B **9**, 600 (1974).

[48] E. O. Kane, in Handbook on Semiconductors Vol. I, edited by W. Paul (North Holland, Amsterdam 1982).

[49] T. B. Bahder, Phys. Rev. B **41**, 11992 (1990).

[50] T. B. Bahder, Phys. Rev. B **45**, 1629 (1992).

[51] P. Enders, phys. stat. sol. (b) **187**, 541 (1995).

[52] P. Enders, A. Bärwolff, M. Woerner, D. Suisky, Phys. Rev. B **51**, 16695 (1995).

[53] S. L. Chuang, Physics of Optoelectronic Devices (John Wiley, New York, 1995).

[54] M. Altarelli, Band Structure, Impurities and Excitons in Superlattices, Heterojunctions and Semiconductor Superlattices, ed. by Allan et al. (Springer, Berlin, 1986).

[55] G. Bastard, Wave Mechanics Applied to Semiconductor Heterostructures (Les Éditions de Physique, Les Ulis, 1988).

[56] G. L. Bir, G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1974).

[57] A. S. Saada, Elasticity: Theory and applications (Robert E. Krieger Publishing Co., Florida, 1989).

[58] C. G. Van de Walle, Phys. Rev. B **39**, 1871 (1989).

[59] G. A. Baraff, D. Gershoni, Phys. Rev. B **43**, 4011 (1991).

[60] M.-E. Pistol, M. Gerling, D. Hessman, L. Samuelson, Phys. Rev. B **45**, 3628 (1992).

[61] Properties of gallium arsenide, second edition (INSPEC, London, 1990).

[62] S. Adachi, Physical properties of III-V semiconductor compounds (John Wiley & Sons, New York, 1992).

[63] G. H. Li, A. R. Goñi, C. Abraham, K. Syassen, P. V. Santos, A. Cantarero, O. Brandt, K. Ploog, Phys. Rev. B **50**, 1575 (1994).

[64] N. Bouarissa, H. Aourag, Mat. Sci. Eng. B **33**, 122 (1995).

[65] S.-H. Wei, A. Zunger, Phys. Rev. B **60**, 5404 (1999).

[66] E. T. Yu, J. O. McCaldin, T. C. McGill, Sol. Stat. Phys. **46**, 1 (1992).

[67] S.-H. Wei, A. Zunger, Appl. Phys. Lett. **72**, 2011 (1998).

[68] S. Adachi (Ed.), Properties of aluminium gallium arsenide (INSPEC, London, 1993).

[69] Properties of lattice-matched and strained indium gallium arsenide, ed. by P. Bhattacharya (INSPEC, London, 1993).

[70] H. Fu, L. Wang, A. Zunger, Appl. Phys. Lett. **73**, 1157 (1998).

[71] L. Wang, A. J. Williamson, A. Zunger, H. Jiang, J. Singh, Appl. Phys. Lett. **76**, 339 (2000).

[72] A. R. Goñi, K. Strössner, K. Syassen, M. Cardona, Phys. Rev. B **36**, 1581 (1987).

[73] A. Zunger (private communication).

[74] K. Hübner, phys. stat. sol. (b) **57**, 627 (1973).

[75] G. R. Zeller, phys. stat. sol. (b) **65**, 521 (1974).

[76] S. de Gironcoli, S. Baroni, R. Resta, Phys. Rev. Lett. **62**, 2853 (1989).

[77] S. de Gironcoli, S. Baroni, R. Resta, Ferroelectrics **111**, 19 (1990).

[78] R. D. King-Smith, D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[79] D. J. Dugdale, S. Brand, R. A. Abram, Phys. Rev. B **61**, 12933 (2000).

[80] Y. Heiner, O. Stier, V. Türck, J. Waschull, B. Sumpf, A. Ostermeier, J. Quant. Spectrosc. Radiat. Transfer **56**, 769 (1996).

[81] A. J. Williamson, A. Zunger, Phys. Rev. B **61**, 1978 (2000).

[82] D. M. Wood, A. Zunger, Phys. Rev. B **53**, 7949 (1996).

[83] L.-W. Wang, A. Zunger, Phys. Rev. B **54**, 11417 (1996).

[84] A. S. Dawydow, Quantenmechanik, 7. Aufl. (VEB Deutscher Verlag der Wissenschaften, Berlin, 1987).

[85] Al. L. Efros, A. L. Efros, Sov. Phys. Semicond. **16**, 772 (1982).

[86] L. Banyai, Y. Z. Hu, M. Lindberg, S. W. Koch, Phys. Rev. B **38**, 8142 (1988).

[87] O. Madelung, Introduction to Solid-State Theory, third edition (Springer Series in Solid-State Sciences, Vol. **2**, Berlin, 1996).

[88] F. Bassani, G. Pastori Parravicini, Electronic States and Optical Transitions in Solids (Pergamon, Oxford, 1975).

[89] A. Barenco, M. A. Dupertuis, Phys. Rev. B **52**, 2766 (1995).

[90] H. Fu, L. Wang, A. Zunger, Appl. Phys. Lett. **71**, 3433 (1997). Comment by Al. L. Efros, M. Rosen, Appl. Phys. Lett. **73**, 1155 (1998). Response by H. Fu, L. Wang, A. Zunger, Appl. Phys. Lett. **73**, 1157 (1998).

[91] M. Grundmann, J. Christen, N. N. Ledentsov, J. Böhrer, D. Bimberg, S. S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V. M. Ustinov, A. Yu. Egorov, A. E. Zhukov, P. S. Kop'ev, Zh. I. Alferov, Phys. Rev. Lett. **74**, 4043 (1995).

[92] F. Heinrichsdorff, MOCVD growth and laser applications of In(Ga)As/GaAs Quantum Dots, Thesis (Mensch & Buch Verlag, Berlin, 1998).

[93] M. Grundmann, O. Stier, S. Bognár, C. Ribbat, F. Heinrichsdorff, D. Bimberg, phys. stat. sol. (a) **178**, 255 (2000).

[94] M. Fricke, A. Lorke, J. P. Kotthaus, G. Medeiros-Ribeiro, P. M. Petroff, Europhys. Lett. **36**, 197 (1996).

[95] C. M. A. Kapteyn, F. Heinrichsdorff, O. Stier, R. Heitz, M. Grundmann, N. D. Zakharov, D. Bimberg, P. Werner, Phys. Rev. B **60**, 14265 (1999).

[96] M.-H. Mao, F. Heinrichsdorff, A. Krost, D. Bimberg, Electron. Lett. **33**, 1641 (1997).

[97] S. Ruvimov, P. Werner, K. Scheerschmidt, J. Heydenreich, U. Richter, N. N. Ledentsov, M. Grundmann, D. Bimberg, V. M. Ustinov, A. Yu. Egorov, P. S. Kop'ev, Zh. I. Alferov, Phys. Rev. B **51**, 14766 (1995).

[98] G. S. Solomon, M. C. Larson, J. S. Harris, Appl. Phys. Lett. **69**, 1897 (1996).

[99] Y. Nabetani, T. Ishikawa, S. Noda, A. Sasaki, J. Appl. Phys. **76**, 347 (1994).

[100] J. M. Moison, F. Houzay, F. Barthe, L. Leprince, E. André, O. Vatel, Appl. Phys. Lett. **64**, 196 (1994).

[101] M. Tabuchi, S. Noda, A. Sasaki, Science and Technology of Mesoscopic Structures, ed. by S. Namba, C. Hamaguchi, T. Ando (Springer, Tokyo, 1992), p. 379.

[102] H. Lee, R. Lowe-Webb, W. Yang, P. Sercel, Appl. Phys. Lett. **72**, 812 (1998).

[103] H. Eisele, O. Flebbe, T. Kalka, C. Preinesberger, F. Heinrichsdorff, A. Krost, D. Bimberg, M. Dähne-Prietsch, Appl. Phys. Lett. **75**, 106 (1999).

[104] E. Dekel, D. Gershoni, E. Ehrenfreund, J. M. Garcia, P. M. Petroff, Proc. 24th International Conference on the Physics of Semiconductors, Jerusalem, Israel, 1998, ed. by D. Gershoni, (World Scientific, Singapore, 1999), No. 1387.

[105] Y. Toda, K. Suzuki, M. Nishioka, M. Moriwaki, Y. Arakawa, Proc. 24th International Conference on the Physics of Semiconductors, Jerusalem, Israel, 1998, ed. by D. Gershoni, (World Scientific, Singapore, 1999), No. 8008.

[106] M. Bayer, A. Kuther, A. Forchel, A. Gorbunov, V. B. Timofeev, F. Schäfer, J. P. Reithmaier, T. L. Reinecke, S. N. Walck, Phys. Rev. Lett. **82**, 1748 (1999).

[107] M. Bayer, A. Kuther, F. Schäfer, J. P. Reithmaier, A. Forchel, Phys. Rev. B **60**, R8481 (1999).

[108] M. Bayer, O. Stern, A. Kuther, A. Forchel, Phys. Rev. B **61**, 7273 (2000).

[109] L. Landin, M.-E. Pistol, C. Pryor, M. Persson, L. Samuelson, M. Miller, Phys. Rev. B **60**, 16640 (1999).

[110] C. Guasch, C. M. Sotomayor Torres, N. N. Ledentsov, D. Bimberg, V. M. Ustinov, P. S. Kop'ev, Superlattices and Microstructures **21**, 509 (1997).

[111] M. J. Steer, D. J. Mowbray, W. R. Tribe, M. S. Skolnick, M. D. Sturge, M. Hopkinson, A. G. Cullis, C. R. Whitehouse, R. Murray, Phys. Rev. B **54**, 17738 (1996).

[112] L. R. Wilson, D. J. Mowbray, M. S. Skolnick, M. Morifuji, M. J. Steer, I. A. Larkin, M. Hopkinson, Phys. Rev. B **57**, R2073 (1998).

[113] I. E. Itskevich, I. A. Trojan, S. G. Lyapin, M. J. Steer, L. R. Wilson, D. Mowbray, M. S. Skolnick, M. Hopkinson, L. Eaves, P. C. Main, Proc. 24th International Conference on the Physics of Semiconductors, Jerusalem, Israel, 1998, ed. by D. Gershoni, (World Scientific, Singapore, 1999), No. 8022.

[114] M. Grundmann, O. Stier, D. Bimberg, Phys. Rev. B **52**, 11969 (1995).

[115] J.-Y. Marzin, G. Bastard, Solid State Commun. **92**, 437 (1994).

[116] V. A. Shchukin, N. N. Ledentsov, V. M. Ustinov, Yu. G. Musikhin, V. B Volovik, A. Schliwa, O. Stier, R. Heitz, D. Bimberg, MRS Symposium Proceedings **618**, 79 (2000).

[117] P. N. Keating, Phys. Rev. **145**, 637 (1966).

[118] R. M. Martin, Phys. Rev. B **1**, 4005 (1970).

[119] M. J. P. Musgrave, J. A. Pople, Proc. Roy. Soc. London **A 268**, 474 (1962).

[120] E. O. Kane, Phys. Rev. B **31**, 7865 (1985).

[121] M. Podgorny, M. T. Czyzyk, A. Balzarotti, P. Letardi, N. Motta, A. Kisiel, M. Zimnal-Starnawska, Solid State Commun. **55**, 413 (1985).

[122] C. Pryor, J. Kim, L.-W. Wang, A. Williamson, A. Zunger, J. Appl. Phys. **83**, 2548 (1998).

[123] H. J. McSkimin, P. Andreatch, J. Appl. Phys. **35**, 3312 (1964).

[124] H. J. McSkimin, P. Andreatch, J. Appl. Phys. **38**, 2610 (1967).

[125] A. J. Williamson, L. W. Wang, A. Zunger, Phys. Rev. B **62**, 12963 (2000).

[126] H. L. McMurry, A. W. Solbrig Jr., J. K. Boyter, C. Noble, J. Phys. Chem. Solids **28**, 2359 (1967).

[127] A. W. Solbrig Jr., J. Phys. Chem. Solids **32**, 1761 (1971).

[128] V. Türck, O. Stier, F. Heinrichsdorff, M. Grundmann, D. Bimberg, Appl. Phys. Lett. **67**, 1712 (1995).

[129] V. Türck, O. Stier, F. Heinrichsdorff, M. Grundmann, D. Bimberg, Phys. Rev. B **55**, 7733 (1997).

[130] L. V. Keldysh, Fizika Tverdogo Tela (Sov. Phys. Solid State) **4**, 2265 (1962).

[131] D. B. Tran Thoai, R. Zimmermann, M. Grundmann, D. Bimberg, Phys. Rev. B **42**, 5906 (1990).

[132] M. Grundmann, O. Stier, D. Bimberg, Phys. Rev. B **58**, 10557 (1998).

[133] T. E. Ostromek, Phys. Rev. B **54**, 14467 (1996).

[134] P. Blood, IEEE J. Quant. Elec. **36**, 354 (2000).

[135] C. Cohen-Tannoudji, J. Dupont-Roc, G. Grynberg, Photons & Atoms (J. Wiley & Sons, New York, 1989).

[136] U. E. H. Laheld, F. B. Pedersen, P. C. Hemmer, Phys. Rev. B **52**, 2697 (1995).

[137] S. V. Nair, T. Takagahara, Phys. Rev. B **55**, 5153 (1997).

[138] M. Braskén, M. Lindberg, D. Sundholm, J. Olsen, Phys. Rev. B **61**, 7652 (2000).

[139] G. W. Bryant, Phys. Rev. Lett. **59**, 1140 (1987).

[140] G. W. Bryant, Phys. Rev. B **37**, 8763 (1988).

[141] Y. Z. Hu, S. W. Koch, M. Lindberg, N. Peyghambarian, E. L. Pollock, F. F. Abraham, Phys. Rev. Lett. **64**, 1805 (1990).

[142] Y. Z. Hu, H. Gießen, N. Peyghambarian, S. W. Koch, Phys. Rev. B **53**, 4814 (1996).

[143] K. Leung, S. Pokrant, K. B. Whaley, Phys. Rev. B **57**, 12291 (1998).

[144] A. Franceschetti, H. Fu, L. W. Wang, A. Zunger, Phys. Rev. B **60**, 1819 (1999).

[145] F. Hatami, M. Grundmann, N. N. Ledentsov, F. Heinrichsdorff, R. Heitz, J. Böhrer, D. Bimberg, S. S. Ruvimov, P. Werner, S. V. Ivanov, B. Ya. Meltser, V. M. Ustinov, P. S. Kop'ev, Zh. I. Alferov, Phys. Rev. B **57**, 4635 (1998).

[146] J. M. Rorison, Phys. Rev. B **48**, 4643 and 17651 (1993).

[147] U. E. H. Laheld, F. B. Pedersen, P. C. Hemmer, Phys. Rev. B **48**, 4659 (1993).

[148] D. Jovanovic, J. Leburton, Phys. Rev. B **49**, 7474 (1994).

[149] M. Stopa, Phys. Rev. B **54**, 13767 (1996).

[150] S. Nagaraja, P. Matagne, V.-Y. Thean, J.-P. Leburton, Y.-H. Kim, R. M. Martin, Phys. Rev. B **56**, 15752 (1997).

[151] L. Landin, M. S. Miller, M.-E. Pistol, C. E. Pryor, L. Samuelson, Science **280**, 262 (1998).

[152] S. Nomura, L. Samuelson, C. Pryor, M.-E. Pistol, M. Stopa, K. Uchida, N. Miura, T. Sugano, Y. Aoyagi, Phys. Rev. B **58**, 6744 (1998).

[153] R. Cingolani, R. Rinaldi, H. Lipsanen, M. Sopanen, R. Virkkala, K. Maijala, J. Tulkki, J. Ahopelto, K. Uchida, N. Miura, Y. Arakawa, Phys. Rev. Lett. **83**, 4832 (1999).

[154] M. Koskinen, M. Manninen, S. M. Reimann, Phys. Rev. Lett. **79**, 1389 (1997).

[155] S. M. Reimann, M. Koskinen, M. Manninen, B. R. Mottelson, Phys. Rev. Lett. **83**, 3270 (1999).

[156] D. G. Austing, S. Sasaki, S. Tarucha, S. M. Reimann, M. Koskinen, M. Manninen, Phys. Rev. B **60**, 11514 (1999).

[157] D. G. Austing, S. Sasaki, S. Tarucha, S. M. Reimann, M. Koskinen, M. Manninen, Physica B **272**, 68 (1999).

[158] I. H. Lee, V. Rao, R. M. Martin, J.-P. Leburton, Phys. Rev. B **57**, 9035 (1998).

[159] S. Nagaraja, L. R. C. Fonseca, J.-P. Leburton, Phys. Rev. B **59**, 14880 (1999).

[160] T. Takagahara, Phys. Rev. B **39**, 10206 (1989).

[161] T. Takagahara, Phys. Rev. B **47**, 4569 (1993).

[162] R. Romestain, G. Fishman, Phys. Rev. B **49**, 1774 (1994).

[163] P. Lelong, G. Bastard, Il Nuovo Cimento **17D**, 1579 (1995).

[164] Ph. Lelong, G. Bastard, Solid State Commun. **98**, 819 (1996).

[165] R. S. Knox, Theory of excitons, Solid State Physics **5** (Academic Press, New York, 1963).

[166] M. Nirmal, D. J. Norris, M. Kuno, M. G. Bawendi, Al. L. Efros, M. Rosen, Phys. Rev. Lett. **75**, 3728 (1995).

[167] Al. L. Efros, M. Rosen, M. Kuno, M. Nirmal, D. J. Norris, M. Bawendi Phys. Rev. B **54**, 4843 (1996).

[168] A. Wojs, P. Hawrylak, Solid State Commun. **100**, 487 (1996).

[169] Ph. Lelong, O. Heller, G. Bastard, Solid-State Electronics **42**, 1251 (1998).

[170] V. D. Kulakovskii, G. Bacher, R. Weigand, T. Kümmell, A. Forchel, E. Borovitskaya, K. Leonardi, D. Hommel, Phys. Rev. Lett. **82**, 1780 (1999).

[171] S. Nomura, Y. Segawa, T. Kobayashi, Phys. Rev. B **49**, 13571 (1994).

[172] D. S. Citrin, Y.-C. Chang, Phys. Rev. B **40**, 5507 (1989).

[173] M. Notomi, S. Nojima, M. Okamoto, H. Iwamura, T. Tamamura, J. Hammersberg, H. Weman, Phys. Rev. B **52**, 11073 (1995).

[174] J. Tulkki, A. Heinämäki, Phys. Rev. B **52**, 8239 (1995).

[175] A. Franceschetti, A. Zunger, Phys. Rev. Lett. **78**, 915 (1997).

[176] S. V. Goupalov, E. L. Ivchenko, J. Crystal Growth **184/185**, 393 (1998).

[177] R. J. Warburton, B. T. Miller, C. S. Dürr, C. Bödefeld, K. Karrai, J. P. Kotthaus, G. Medeiros-Ribeiro, P. M. Petroff, S. Huant, Phys. Rev. B **58**, 16221 (1998).

[178] H. De Raedt, M. Frick, Physics Reports **231**, 107 (1993)

[179] A. J. Williamson, A. Franceschetti, A. Zunger, Europhysics Letters, to be published.

[180] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, Numerical Recipes in C (Cambridge University Press, 1992).

[181] Y. Z. Hu, M. Lindberg, S. W. Koch, Phys. Rev. B **42**, 1713 (1990).

[182] E. L. Ivchenko, phys. stat. sol. (a) **164**, 487 (1997).

[183] S. Yokojima, T. Meier, S. Mukamel, J. Chem. Phys. **106**, 3837 (1996).

[184] H. Fu, L. W. Wang, A. Zunger, Phys. Rev. B **59**, 5568 (1999).

[185] A. Hartmann, Y. Ducommun, E. Kapon, U. Hohenester, E. Molinari, Phys. Rev. Lett. **84**, 5648 (2000).

[186] M. Grundmann, D. Bimberg, Phys. Rev. B **55**, 9740 (1997).

[187] M. Grundmann, D. Bimberg, Jpn. J. Appl. Phys. **36**, 4181 (1997).

[188] S. Sauvage, P. Boucaud, J. -M. Gérard, V. Thierry-Mieg, Phys. Rev. B **58**, 10562 (1998).

[189] S. J. Chua, S. J. Xu, X. H. Zhang, X. C. Wang, T. Mei, W. J. Fan, C. H. Wang, J. Jiang, X. G. Xie, Appl. Phys. Lett. **73**, 1997 (1998).

[190] J. Phillips, P. Bhattacharya, S. W. Kennerly, D. W. Beekman, M. Dutta, IEEE J. Quant. Elec. **35**, 936 (1999).

[191] G. D. Mahan, Many-Particle Physics (Plenum, New York, 1981).

[192] K. Huang, A. Rhys, Proc. R. Soc. London, Ser. A **204**, 406 (1950).

[193] S. Nomura, T. Kobayashi, Phys. Rev. B **45**, 1305 (1992).

[194] S. Rudin, T. L. Reinecke, Phys. Rev. B **41**, 3017 (1990).

[195] R. Heitz, I. Mukhametzanov, O. Stier, A. Madhukar, D. Bimberg, Phys. Rev. Lett. **83**, 4654 (1999).

[196] R. Heitz, I. Mukhametzhanov, O. Stier, A. Madhukar, D. Bimberg, Physica E **7**, 398 (2000).

[197] A. Garcia-Cristóbal, A. W. E. Minnaert, V. M. Fomin, J. T. Devreese, A. Yu. Silov, J. E. M. Haverkort, J. H. Wolter, phys. stat. sol. (b) **215**, 331 (1999).

[198] H. Lipsanen, M. Sopanen, J. Ahopelto, Phys. Rev. B **51**, 13868 (1995).

[199] M. Grundmann, N. N. Ledentsov, O. Stier, J. Böhrer, D. Bimberg, V. M. Ustinov, P. S. Kop'ev, Zh. I. Alferov, Phys. Rev. B **53**, R10509 (1996).

[200] S. Noda, T. Abe, M. Tamura, Physica E **2**, 643 (1998).

[201] U. Bockelmann, W. Heller, A. Filoramo, Ph. Roussignol, G. Abstreiter, phys. stat. sol. (a) **164**, 281 (1997).

[202] G. Guttroff, M. Bayer, A. Forchel, D. V. Kazantsev, M. K. Zundel, K. Eberl, phys. stat. sol. (a) **164**, 291 (1997).

[203] M. Markmann, A. Zrenner, G. Böhm, G. Abstreiter, phys. stat. sol. (a) **164**, 301 (1997).

[204] K. Brunner, G. Abstreiter, G. Böhm, G. Tränkle, G. Weimann, Phys. Rev. Lett. **73**, 1138 (1994).

[205] M. Notomi, T. Furuta, H. Kamada, J. Temmyo, T. Tamamura, Phys. Rev. B **53**, 15743 (1996).

[206] D. Gammon, E. S. Snow, B. V. Shanabrook, D. S. Katzer, D. Park, Science **273**, 87 (1996).

[207] D. Hessman, P. Castrillo, M.-E. Pistol, C. Pryor, L. Samuelson, Appl. Phys. Lett. **69**, 749 (1996).

[208] A. Gustafsson, M.-E. Pistol, L. Montelius, L. Samuelson, J. Appl. Phys. **84**, 1715 (1998).

[209] Y. Toda, S. Shinomori, K. Suzuki, Y. Arakawa, Phys. Rev. B **58**, R10147 (1998).

[210] A. Zrenner, M. Markmann, E. Beham, F. Findeis, G. Böhm, G. Abstreiter, J. Electr. Materials **28**, 542 (1999).

[211] R. Heitz, A. Kalburge, Q. Xie, M. Grundmann, P. Chen, A. Hoffmann, A. Madhukar, D. Bimberg, Phys. Rev. B **57**, 9050 (1998).

[212] Q. Xie, P. Chen, A. Kalburge, T. R. Ramachandran, A. Nayfonov, A. Konkar, A. Madhukar, J. Crys. Growth **150**, 357 (1995).

[213] H. Benisty, C. M. Sotomayor-Torres, C. Weisbuch, Phys. Rev. B **44**, 10945 (1991).

[214] T. Inoshita, H. Sakaki, Phys. Rev. B **46**, 7260 (1992).

[215] R. Heitz, M. Veit, N. N. Ledentsov, A. Hoffmann, D. Bimberg, V. M. Ustinov, P. S. Kop'ev, Zh. I. Alferov, Phys. Rev. B **56**, 10435 (1997).

[216] I. Mukhametzhanov, R. Heitz, J. Zeng, P. Chen, A. Madhukar, Appl. Phys. Lett. **73**, 1341 (1998).

[217] R. Heitz, I. Mukhametzhanov, P. Chen, A. Madhukar, Phys. Rev. B **58**, R10151 (1998).

[218] I. Mukhametzhanov, J. Wei, R. Heitz, A. Madhukar, Appl. Phys. Lett. **75**, 85 (1999).

[219] F. Heinrichsdorff, M. Grundmann, O. Stier, A. Krost, D. Bimberg, J. Cryst. Growth **195**, 540 (1998).

[220] U. Banin, C. J. Lee, A. A. Guzelian, A. V. Kadavanich, A. P. Alivisatos, W. Jaskolski, G. W. Bryant, Al. L. Efros, M. Rosen, J. Chem. Phys. **109**, 2306 (1998).

[221] C. M. A. Kapteyn, M. Lion, R. Heitz, D. Bimberg, P. Brunkov, B. V. Volovik, S. G. Konnikov, A. R. Kovsh, V. M. Ustinov, Appl. Phys. Lett. **76**, 1573 (2000).

[222] S. Tiwari, G. D. Pettit, K. R. Milkove, F. Legoues, R. J. Davis, J. M. Woodall, Appl. Phys. Lett. **64**, 3536 (1994).

[223] E. Kapon, D. M. Hwang, M. Walther, R. Bhat, N. G. Stoffel, Surf. Sci. **267**, 593 (1992).

[224] F. Vouilloz, D. Y. Oberli, M.-A. Dupertuis, A. Gustafsson, F. Reinhardt, E. Kapon, Phys. Rev. B **57**, 12378 (1998).

[225] O. Stier, V. Türck, M. Kappelt, D. Bimberg, Physica E **2**, 969 (1998).

[226] M. Kappelt, V. Türck, O. Stier, D. Bimberg, D. Stenkamp, Proc. 9th International Conference on Indium Phosphide and Related Materials, Hyannis, USA, IEEE, p. 83 (1997).

[227] F. Lelarge, C. Constantin, K. Leifer, A. Condo, V. Iakovlev, E. Martinet, A. Rudra, E. Kapon, Appl. Phys. Lett. **75**, 3300 (1999).

[228] E. L. Martinet, M.-A. Dupertuis, E. Kapon, O. Stier, M. Grundmann, D. Bimberg, Proc. 24th International Conference on the Physics of Semiconductors, Jerusalem, Israel, 1998, ed. by D. Gershoni, (World Scientific, Singapore, 1999), No. 1449.

[229] Y. Arakawa, H. Sakaki, Appl. Phys. Lett. **40**, 939 (1982).

[230] C. R. Proetto, Phys. Rev. B **45**, 11911 (1992).

[231] T. Tadić, Z. Ikonić, Phys. Rev. B **50**, 7680 (1994).

[232] G. Bastard, J. Y. Marzin, Solid State Commun. **91**, 39 (1994).

[233] M. Bayer, C. Schlier, C. Gréus, A. Forchel, S. Benner, H. Haug, Jpn. J. Appl. Phys. **34**, 4408 (1995).

[234] M. Grundmann, J. Christen, M. Joschko, D. Bimberg, E. Kapon, Proc. 22nd International Conference on the Physics of Semiconductors, Vancouver, Canada, 1994, ed. by D. J. Lockwood (World Scientific, Singapore, 1995), 1675.

[235] R. Ambigapathy, I. Bar-Joseph, D. Y. Oberli, S. Haacke, M. J. Brasil, F. Reinhardt, E. Kapon, B. Deveaud, Phys. Rev. Lett. **78**, 3579 (1997).

[236] C. R. Bennett, K. Güven, B. Tanatar, Phys. Rev. B **57**, 3994 (1998).

[237] E. H. Hwang, S. Das Sarma, Phys. Rev. B **58**, R1738 (1998).

[238] F. Vouilloz, S. Wiesendanger, D. Y. Oberli, B. Dwir, F. Reinhardt, E. Kapon, Physica E **2**, 862 (1998).

[239] M. Grundmann, O. Stier, D. Bimberg, Phys. Rev. B **50**, 14187 (1994).

[240] O. Mauritz, U. Ekenberg, Proc. 23rd International Conference on the Physics of Semiconductors, Berlin, Germany, 1996, ed. by M. Scheffler and R. Zimmermann (World Scientific, Singapore, 1996), p. 1823.

[241] F. Rossi, E. Molinari, Phys. Rev. Lett. **76**, 3642 (1996).

[242] M. Grundmann, O. Stier, A. Schliwa, D. Bimberg, Phys. Rev. B **61**, 1744 (2000).

[243] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).

[244] J. C. Yi, N. Dagli, IEEE J. Quant. Elec. **31**, 208 (1995).

[245] M. G. Feler, J. Comp. Phys. **14**, 341 (1974).

[246] L. Wang, A. Zunger, J. Chem. Phys. **100**, 2394 (1994).

[247] O. Stier, Elektronische Eigenschaften pseudomorpher Quantenfäden (Master thesis, Technische Universität Berlin, Institut für Festkörperphysik, 1995).

[248] Y. Arakawa, Y. Yamauchi, J. N. Schulman, Phys. Rev. B **43**, 4732 (1991).

[249] P. Vogl, H. P. Hjalmarson, J. D. Dow, J. Phys. Chem. Solids **44**, 365 (1983).

[250] T. Yamauchi, Y. Arakawa, J. N. Schulman, Appl. Phys. Lett. **57**, 1224 (1990).

[251] T. Yamauchi, T. Takahashi, Y. Arakawa, IEEE J. Quant. Elec. **27**, 1817 (1991).

[252] M. T. Michalewicz, Comp. Phys. Comm. **79**, 13 (1994).

[253] M. Dib, M. Chamarro, V. Voliotis, J. L. Fave, C. Guenaud, P. Roussignol, T. Gacoin, J. P. Boilot, C. Delerue, G. Allan, M. Lannoo, phys. stat. sol. (b) **212**, 293 (1999).

[254] K. Dowd, High Performance Computing (O'Reilly & Assoc., Sebastopol, CA, 1993)

[255] M. Pistol, D. Gershoni, Phys. Rev. B **50**, 11738 (1994).

[256] L. Pfeiffer, H. Baranger, D. Gershoni, K. Smith, W. Wegscheider, NATO ASI Series E Vol. **298**, p. 93 (Kluwer, Dordrecht, 1995).

[257] K. Kojima, K. Mitsunaga, K. Kyuma, Appl. Phys. Lett. **55**, 882 (1989).

[258] H. K. Harbury, W. Porod, J. Vac. Sci. Technol. B **8**, 923 (1990).

[259] H. K. Harbury, W. Porod, S. M. Goodnick, J. Appl. Phys. **73**, 1509 (1992).

[260] Z. Wu, P. P. Ruden, J. Appl. Phys. **74**, 6234 (1993).

[261] L. R. Ram-Mohan, J. R. Meyer, J. Nonlinear Optical Phys. and Materials **4**, 191 (1995).

[262] I. Vurgaftman, J. M. Hinckley, J. Singh, IEEE J. Quant. Elec. **30**, 75 (1994).

[263] S. E. Laux, F. Stern, Appl. Phys. Lett. **49**, 91 (1986).

[264] T. Kerkhoven, A. T. Gallick, U. Ravaioli, J. H. Arends, Y. Saad, J. Appl. Phys. **68**, 3461 (1990).

[265] A. Kumar, S. E. Laux, F. Stern, Phys. Rev. B **42**, 5166 (1990).

[266] C. Pryor, Phys. Rev. B **44**, 12912 (1991).

[267] A. Nakano, P. Vashishta, R. K. Kalia, Comp. Phys. Comm. **83**, 181 (1994).

[268] D. S. Citrin, Y. Chang, Optical Materials **1**, 49 (1992); IEEE J. Quant. Elec. **29**, 97 (1993).

[269] D. S. Citrin, Y. Chang, J. Appl. Phys. **68**, 161 (1990).

[270] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[271] L. Wang, A. Zunger, *Pseudopotential Theory of Nanometer Silicon Quantum Dots*, in: P. V. Kamat, D. Meisel (Eds.), Nanocrystalline Semiconductor Materials (Elsevier Science, 1996).

[272] C. H. Park, I. Lee, K. J. Chang, Phys. Rev. B **47**, 15996 (1993).

[273] R. A. Lippert, M. A. Sears, Phys. Rev. B **61**, 12772 (2000).

[274] B. N. Parlett, The Symmetric Eigenvalue Problem (Prentice-Hall, Englewood Cliffs, 1980).

[275] J. K. Cullum, R. A. Willoughby, Lanczos Algorithms for Large Sparse Eigenvalue Computations Vol. I, Theory (Birkhäuser, 1985).

[276] N. S. Sehmi, Large Order Structural Eigenanalysis Techniques (Ellis Horwood Ltd., Chichester, 1989).

[277] Y. Saad, Numerical methods for large eigenvalue problems (Manchester University Press, 1992).

[278] D. M. Wood, A. Zunger, J. Phys. A **18**, 1343 (1985).

[279] H. R. Schwarz, Lecture Notes in Mathematics Vol. **968**, p. 384 (Springer, Berlin, 1982).

[280] H. Tal-Ezer, R. Kosloff, C. Cerjan, J. Comp. Phys. **100**, 179 (1992).

[281] R. Kosloff, H. Tal-Ezer, Chem. Phys. Lett. **127**, 223 (1986).

[282] R. G. Grimes, J. G. Lewis, H. D. Simon, SIAM J. Matrix Anal. Appl. **15**, 228 (1994).

[283] G. Grosso, L. Martinelli, G. Pastori Parravicini, Phys. Rev. B **51**, 13033 (1995).

[284] A. Meyer, Modern Algorithms for Large Sparse Eigenvalue Problems (Akademie-Verlag, Berlin, 1987).

[285] H. Yang, Computing **51**, 79 (1993).

[286] G. Gambolati, G. Pini, M. Putti, SIAM J. Sci. Comput. **16**, 173 (1995).

[287] O. Axelsson, Iterative Solution Methods (Cambridge University Press, 1996).

[288] H. G. Matthies, Comp. Struc. **21**, 319 (1985).

[289] S. J. Sciutto, Comp. Phys. Comm. **77**, 95 (1993).

[290] L. Verlet, Phys. Rev. **159**, 98 (1967).

[291] Y. Okamoto, H. J. Maris, Comp. Phys. Comm. **76**, 191 (1993).

[292] S. J. Sciutto, Comp. Phys. Comm. **79**, 215 (1994).

[293] M. C. Payne, J. D. Joannopoulos, D. C. Allan, M. P. Teter, D. H. Vanderbilt, Phys. Rev. Lett. **56**, 2656 (1986).

[294] A. Kiełbasiński, H. Schwetlick, Numerische lineare Algebra (VEB Deutscher Verlag der Wissenschaften, Berlin, 1988).

[295] R. B. Morgan, D. S. Scott, SIAM J. Sci. Comput. **14**, 585 (1993).

[296] A. H. Sameh, J. A. Wisniewski, SIAM J. Numer. Anal. **19**, 1243 (1982).

[297] F. A. Dul, K. Arczewski, J. Comp. Phys. **111**, 89 (1994).

[298] R. B. Morgan, D. S. Scott, SIAM J. Sci. Stat. Comput. **7**, 817 (1986).

[299] R. Bru, C. Corral, A. Martinez, J. Mas, SIAM J. Matrix Anal. Appl. **16**, 1210 (1995).

[300] D. J. Evans, J. Shanehchi, Comp. Meth. Appl. Mech. Eng. **31**, 251 (1982).

[301] J. L. Martins, M. L. Cohen, Phys. Rev. B **37**, 6134 (1988).

[302] R. B. Morgan, J. Comp. Phys. **89**, 241 (1990).

[303] R. B. Morgan, J. Comp. Phys. **101**, 287 (1992).

[304] E. R. Davidson, J. Comp. Phys. **17**, 87 (1975).

[305] W. Butscher, W. E. Kammer, J. Comp. Phys. **20**, 313 (1976).

[306] C. W. Murray, S. C. Racine, E. R. Davidson, J. Comp. Phys. **103**, 382 (1992).

[307] E. R. Davidson, Comp. Phys. **7**, 519 (1993).

[308] A. Stathopoulos, C. F. Fischer, Comp. Phys. Comm. **79**, 268 (1994).

[309] E. R. Davidson, J. Phys. A **13**, L179 (1980).

[310] N. Kosugi, J. Comp. Phys. **55**, 426 (1984).

[311] L. Borges, S. Oliveira, J. Comp. Phys. **144**, 727 (1998).

[312] T. Sommerfeld, F. Tarantelli, J. Chem. Phys. **112**, 2106 (2000).

[313] T. P. Hamilton, P. Pulay, J. Chem. Phys. **84**, 5728 (1986).

[314] H. Kim, B. D. Yu, J. Ihm, J. Phys. A **27**, 1343 (1994).

[315] J. L. Martins, N. Troullier, Phys. Rev. B **43**, 2213 (1991).

[316] V. Dolcher, G. Grosso, L. Martinelli, G. Pastori Parravicini, Phys. Rev. B **53**, 10813 (1996).

[317] E. Anderson, J. Brooks, C. Grassl, S. Scott, Performance of the CRAY T3E Multiprocessor, WWW document, Cray Research Inc.

[318] E. Anderson, J. Brooks, T. Hewitt, The Benchmarkers's Guide to Single-Processor Optimization for CRAY T3E Systems, WWW document, Cray Research Inc.

Index

0D . . . 3
1D . . . 3
2D . . . 98
3D . . . 35

absorption coefficient . . . 58
absorption cross section . . . 56
accumulation of information 125, 130, 131
acquisition of information . . . 126, 146
adiabatic . . . 91
alloy . . . 21
angular momentum . . . 37, 59, 67
angular momentum conservation . . . 36
angular quantum number . . . 36
anisotropy . . . 65, 113
antimetrization . . . 67
Arnoldi procedure . . . 130, 142, 150
ASA . . . 21
atomic structure anisotropy . . . 21, 25
average valence band energy . . . 9

band alignment . . . 47, 54, 62
band mixing . . . 3, 37, 50, 52, 59, 67, 79
band offset . . . 10, 13
band-gap renormalization . . . 104
basis function . . . 62
Bauer's method . . . 129
benchmark . . . 159
Bessel function . . . 36
biexciton . . . 65, 81
Bloch function . . . 7, 17
Bloch theorem . . . 8
block Davidson algorithm . . . 146, 160
Bohr radius . . . 35
bond bending/stretching . . . 41, 42
bound state . . . 47
boundary condition . . . 39
Bravais lattice . . . 6
Brillouin zone . . . 6
broadening . . . 93, 104
BZ . . . 11

cache performance . . . 165
calorimetric absorption . . . 38
capacitance spectroscopy . . . 100
carrier temperature . . . 104
cartesian product . . . 66, 70
cathodoluminescence . . . 37
Cayley form . . . 128
CB . . . 7
center-of-mass motion . . . 35
CG . . . 124
charge . . . 65, 68, 80, 104
Chebyshev accelerated power method 128
Chebyshev iteration . . . 125, 132
Chebyshev polynomial . . . 128, 134
CI . . . 62, 125, 132
cleaved-edge overgrowth . . . iv, 114
CM . . . 39
colloidal dots . . . 100
comparison IGDA vs. FSM . . . 159
complex conjugation . . . 8
complexity . . . 126, 149, 151, 154, 162
configuration interaction . . . 62, 66
configuration mixing . . . 62
confinement . . . iv, 35, 37, 88
conjugate . . . 131
conjugate-gradient method . . . 39, 124, 131, 159
conservation . . . 35–37, 59, 76
consistency range . . . 149
consistent . . . 57, 139, 149
constraint . . . 135
continuum mechanical strain model . . . 39, 47, 55
convergence factor . . . 126, 162
convergence rate 126, 133, 134, 150, 154
coordinate relaxation method . . . 125
correction vector . . . 147, 149, 150
correlation . . . 61, 64, 81, 110
Coulomb charging energy . . . 65
Coulomb interaction . . . 61
Coulomb matrix element . . . 23, 28
Coulomb term . . . 68
Cray T3E . . . 164
CRI . . . 162
cut-off . . . 119, 127

DA . . . 146
dark exciton . . . 73
Davidson algorithm . . . 146
decoupled bandstructure models . . . 66, 68, 120
deep level transient spectroscopy . . . 100

deflection of eigenvectors 138
deformation potential 10, 18
degeneracy 59, 63, 70, 79, 154, 157
density functional theory 119
density of states 104
DFT 119
diagonal dominance 145
dielectric constant 46, 68
dielectric function 69
dielectric screening 68
difference quotient 39, 160
DIIS 151
dipole 45
dipole transition 56
direction set method 125
Dirichlet boundary conditions 46
discretization 47, 53, 119, 122
discretization error 53, 136, 139
displacement 39
dissociation 81
domain decomposition 164
DOS 5, 104
double precision 148

EBOM 123
economization 69
EFA 119
effective mass approximation 35, 38, 52, 66, 68, 111, 119
efficiency 142
eigenproblem 120
eigenvalue economization 124
elastic moduli 39, 42
electron picture 72
elimination of redundancy 148
empirical pseudopotential method 5, 17
envelope function 8, 47, 50
envelope function approximation 57
EOM 135
EP 5, 17
epitaxy 37, 94
equation of motion 135
equivalence transformation 122
Euler method 141
exact arithmetic 125, 155
exchange 61, 63, 66
exchange splitting 78
exchange strength constant 68
exchange term 68
exchange-correlation potential 66
exciton binding energy 35, 64
exciton ground state 63
exciton phonon interaction iii, 91
exponential spectral transformation 128
extreme eigenvalues 119, 123

fast Fourier transformation 123
FDM 117, 122
FEM 122
Fermi integral 104
Fermi level 85
Fermion 66
finite barrier 37
finite-differences method 39, 47
finite-precision arithmetic 125, 131
fix vector 132, 136
Fletcher-Reeves 131
floating point operations 122
flops 164
fluctuations 38
FM 122
folded spectrum method 154
four-band model 68, 119
Fröhlich constant 91
Frenkel exciton 68
FSM 154
full orthogonalization 148
full reorthogonalization 132
fundamental band gap 10

gain of information 126, 147, 148
Γ-point 5, 17
gap ratio 126, 142
GDA 125
generalized Davidson algorithm 125, 144, 146
generalized eigenproblem 122
global convergence 150, 151, 153
gradient method 125, 132, 141
Gram-Schmidt procedure 148
GRINSCH 105

harmonic oscillator 96
Hartree approximation 62, 63, 70
Hartree-Fock approximation 72
HH 24
Huang-Rhys parameter 91

IEOM 129
IGDA 154

ill-conditioned 150
image charge effect 46, 63, 69, 71
implicit deflation 146, 152, 156
incomplete factorization 144
independent-phonon model 91
inhomogeneity 94
inhomogeneous line broadening 38
inhomogeneous strain 3
initialisation 146
integer operations 122
interdiffusion 98
interface dipole 13
interface orientation 11, 13
interior eigenvalue 120, 144, 152
interior eigenvector 156
intermediate confinement regime . . . 35
inverse iteration 144
inversion asymmetry 22
iterative extraction orthogonalization method 129

Jacobi method 145

Kane model 8
k conservation 35
kinetic energy 21
Kohn-Sham equations 119
k·**p** approximations 17
k-representation 122
Krylov method 130, 155
Krylov sequence 131
Krylov space 131, 150, 155

LA 125
Lagrange multiplier 135
Lanczos algorithm 125, 131
Lanczos phenomenon 132
laser 38
lattice mismatch 37
LDA 63
LH 24
linear convergence 126
linearized valence force field model . . 42
LO 91
local density approximation 63
local spin density approximation . . . 63
long-range exchange 66, 68
longitudinal effective mass 104
LSDA 63
Luttinger parameter 10

machine precision 148
magnetic field 63, 67
many-body expansion 62
many-particle effects 104
master equations 85
matching condition 17
material parameter 47, 53
matrix factorization 124
matrix inversion 124
matrix transpose 124
matrix vector multiplication 122, 123, 124
metric tensor 122
microstate 85
mid-infrared 58, 90
MIMD 164
min-max property 128, 135
minimization method 124
ML 18
model-solid theory 9
modified Jacobi operator 145
molecular dynamics method 132, 135, 141, 149
momentum conservation 35
momentum matrix element 56, 72
momentum operator 56
MPI 164
MSBN interaction 42
multi-term recurrence 125, 131
multigrid technique 124
multiplicity 156

narrow-gap semiconductor 37, 120
nested iteration 71
nodal surface 24, 50
non-analytic function 138, 143
non-linear strain dependence of band edges 20
non-planar substrate 103
non-unitary basis 122
normalization 137

OC 122
operator polynomial 133
optical matrix element 57
optical matrix parameter 54, 58
orbital function 121
orthonormal basis 68
oscillator strength 56, 58, 78, 86
overconfinement 27

overlap integral 58

Padé approximant 145
parallelization 164
parameter fit 21
parametrization 66
parity 24, 37, 51
partial differential equation 47
PCG 124
performance 126
perturbational effective mass approach 38
phonon 42, 91, 94
photoluminescence excitation 38
piezoelectric effect 20, 45, 54, 83
piezoelectric modulus 45
PL 38
PL spectroscopy 93
plane wave expansion 122
PLE 38
PLE spectroscopy 94
PM 126
Poisson equation 46, 70, 71
Polak-Ribiere 132, 154
polarization 85, 113
polarization anisotropy 28
polarization charge 45
population 61, 85, 93, 104
positive definite 155
potential energy 21
power method 126
preconditioned Chebyshev iteration 139
preconditioned conjugate gradient
    method 124, 143, 154
preconditioned Lanczos algorithm . 144
preconditioning 143, 149
preconditioning matrix . . 139, 141, 144
probability density 24, 50
projection 122
projector 142
pseudopotential calculation 3, 145
pseudopotential method 123
pseudoresidual 147, 153

QD 3
quadrupole potential 83
quantum Monte-Carlo 64, 70
quantum number 37, 50, 59, 76, 96, 103
quantum size effect 37, 38, 95
QWR 3

radial quantum number 36
RAM 162
random vector 107
rank 127
rate equation 85
Rayleigh quotient . . . .
    . . . . . 124, 126, 129, 141, 151, 158
Rayleigh quotient iteration 144
Rayleigh-Ritz procedure . . 71, 147, 129
reciprocal lattice 6
remote band effects 120
renormalization 61
reorthogonalization 132
representation 117, 121, 144, 145
residual 127
residual minimization 129, 157
residual norm 124, 129, 151
restarting 132
Ritz matrix 147
Ritz vector 147
RM 129
RM matrix 157
RM vector 157
RMM-DIIS method 154
round-off error 150
RR 129

scalar product 148
search energy 156
seed layer 94
selection rule 67
selective reorthogonalization 148
self-consistent calculation . . 63, 64, 104
self-consistent mean field 63, 64
self-termination 132, 136, 138, 140, 149
semiorthogonalization 148
separable wave function 63, 64
shift-and-invert 144, 146
shifting strategy 147, 149
short-range exchange 66, 68
simultaneous iteration 125, 129
single-band model 66, 68, 111, 119
single-precision 148
single-QD spectroscopy 94
singlet 76
singlet-triplet splitting 76
six-band model 68, 119
Slater determinant 67
SM 123, 144

SME 58
SO 7
software 162
Solbrigs model 42
SOR 144
sparse 121, 162
spatial charge density 46
spatial representation 47
spatial resolution 122
spectral gap 155
spectral method 123
spectral transformation 127
spin 66
spin conservation 59
spin orbit energy 10
spin orbit interaction 7, 59, 67, 76
spontaneous emission 58
spread 126, 160
SSOR 144
stability condition 136, 137
stagnation of the GDA 153, 154
state filling 93
steepest-descent method 141
stochastic diagonalization 70
Stoermer's rule 136
storage of information 148
strain 10, 39, 54
strain field symmetry 44
strain inhomogeneity 23
Stranski-Krastanow growth 38
strong confinement 35, 61
subband 104
subspace iteration 125, 129
surface charge density 46
symmetric positive definite 146
symmetrization 11
symmetry breaking . iv, 52, 86, 111, 114
symmetry effect 46
symmetry reduction 21

tetrahedral configuration 44
three-term recurrence 125, 131
tight-binding model 121
time reversal symmetry 35, 59, 63, 70, 154, 156
total angular momentum 76
total energy functional 123
total strain energy 39
transition matrix element 23
transmission electron microscopy 105, 114
Treppeniteration 129
trion 65, 80
triplet 76
truncation 38, 132, 147
two-electron ground state 65
type-II 62

unharmonic Keating model 42
unit cell 6
unit stride 165
use of information 129, 147

V-groove iv
valence force field model 19, 39, 55
variable deposition amount 94
variance minimisation 129
variational configuration interaction method 62
variational method 63
VB 5
vector iteration 126
Verlet algorithm 136
VFF 39

Wannier exciton 66, 68
wave-function product 59, 70, 71
weak confinement 35
weak correlation 62, 72
wide-gap semiconductor 120
WL 18

X 69
X-ray photospectroscopy 5
XX 69

ZIB 162