

Unterschrift des Betreuers



DIPLOMARBEIT

Numerical Solution of the Boltzmann Transport Equation using Spherical Harmonics Expansions

Ausgeführt am

Institut für Analysis und Scientific Computing
der Technischen Universität Wien

unter der Anleitung von
Univ.Prof. Dr.rer.nat. Ansgar Jüngel

durch

Karl Rupp

Eslarngasse 27/19
1030 Wien

Matrikelnummer 0325941
Studienkennzahl 864

Datum

Unterschrift

Abstract

This thesis deals with the numerical solution of the Boltzmann transport equation using an expansion of the distribution function in spherical harmonics for the purpose of electronic device simulation. Both the mathematical and physical backgrounds are discussed, then the Boltzmann transport equation is projected onto spherical harmonics without posing unnecessary restrictions on the energy band structure. From entropy principles a stabilisation is found which serves as a Scharfetter-Gummel-like stabilisation for the discretisation. The finite volume method using a full Galerkin scheme is proposed for the discretisation of the projected equations, which has the advantage of ensuring current continuity by virtue of construction. To reduce computational costs and speed up the assembly of the system matrix, analytical formulae for the integral terms in the discretised equations are derived.

Complexity analysis shows that higher order spherical harmonics expansions suffer from huge memory requirements, especially for two and three dimensional devices. A compressed matrix storage scheme is therefore introduced, which reduces the memory requirements for the storage of the system matrix especially for higher order spherical expansions by up to several orders of magnitude.

Finally, simulation results for a n^+nn^+ -diode prove the applicability of the full Galerkin method. Self-consistent solutions are obtained by coupling the system of projected equations with the Poisson equation. The resulting systems of linear equations turn out to be poorly conditioned, thus preconditioners are proposed and compared.

Contents

Introduction	1
1 Spherical Harmonics	3
1.1 The Laplace Equation in Polar Coordinates	3
1.2 Legendre Polynomials	5
1.3 Associated Legendre Functions	7
1.4 Orthogonal Functions on the Sphere	12
1.5 The Spherical Fourier Transform	17
2 The Boltzmann Transport Equation	22
2.1 Derivation of the Transport Equation	22
2.2 Collision Events	26
2.3 Multi-Band Structures	30
2.4 Average Quantities and Moments	32
3 Spherical Projection of the Boltzmann Transport Equation	36
3.1 The Spherical Harmonics Expansion Equations	36
3.2 Parity Splitting	41
3.3 Entropy Considerations	44
4 Discretisation	49
4.1 The Finite Volume Method	49
4.2 The Scharfetter-Gummel Discretisation	52
4.3 Discretisation of the SHE of the BTE	54
4.4 Rectangular Grids in Two Dimensions	58
4.5 Boundary Conditions	64
5 Reduction of Memory Requirements	66
5.1 The Naive Approach	66
5.2 Sparsity of Coupling Matrices	68
5.3 Matrix Factorisation	78
6 Implementation and Results	88
6.1 Problem Setup	88
6.2 Pitfalls	89
6.3 Results for a n^+nn^+ Diode	96

7 Outlook and Conclusion	102
7.1 Unstructured and Adaptive Grids	102
7.2 Adaptive Control of the Expansion Order	103
7.3 Higher Spatial Dimensions	104
7.4 Mathematical Modelling	105
7.5 Appropriate Linear Solver	106
7.6 Finite Element Methods	107
7.7 Conclusion	107
Bibliography	109

List of Figures

1.1	Selected spherical harmonics.	13
2.1	Illustration of trajectories in position-momentum space	27
2.2	Multi-band structure of silicon.	31
4.1	Two methods for the construction of a dual box tessellation from a triangulation.	51
4.2	Labels for the Finite Volume Method	53
4.3	The Bernoulli function $\text{Bern}(x)$	55
4.4	Labels associated with a rectangular box $B_{i,j}$ centered at (x_i, ε_j)	59
6.1	Connection of an n^+nn^+ -diode with a MOSFET in inversion	89
6.2	Eigenvalues of the system matrix with different preconditioners.	95
6.3	Comparison of eigenvalues of the system matrix without and with elimination of odd order unknowns.	96
6.4	Self-consistent solution of SHE equations at an applied voltage of 0.3 V. . .	97
6.5	Self-consistent solution of SHE equations at an applied voltage of 1.0 V. . .	98
6.6	Norm of the potential update vector $\ \Delta\psi_s\ $ after each Gummel iteration step for different externally applied voltages.	99
6.7	The potential and its correction at different Gummel iteration steps.	100
6.8	Coarse grid solutions for a n^+nn^+ diode at an applied voltage of 1.0 V. . .	100
6.9	Comparison between electron concentrations obtained from self consistent, non self consistent and drift-diffusion simulations.	101
7.1	Construction of a prismatic mesh from a lower-dimensional one.	104

List of Tables

2.1	Parameters for the analytical band structure.	32
6.1	Input parameters for the simulation of an n^+nn^+ diode.	90
6.2	Comparison of condition numbers with different preconditioners.	94

Acknowledgement

Many thanks go to Prof. Ansgar Jüngel for his cooperativeness with the interdisciplinary topic of this thesis. He has always been eager to hear about the latest progress of my work and provided me with the highly appreciated freedom that allowed me – driven by my own curiosity – to find the analytical and numerical results.

I am also indebted to Prof. Klaus-Tibor Grasser from the Institute for Microelectronics. I remember when he suggested my investigation of the spherical harmonics expansion within this thesis, then raised his eyebrows and jokingly said that a scientific hybrid like me would definitely enjoy this topic. Yes, he was right.

A hearty thanks to Philipp Dörsek, who provided the cover page of this thesis, and Zlatan Stanojevic, who helped me with the interpretation of the simulation results from the physics point of view.

During my study I have learnt to know many course mates that are now good friends. Many thanks for making my studies even more pleasurable and for spending a lot of spare-time together.

I also wish to thank my parents, grandparents and sisters for the support during my study, even though I have hardly found the time to travel back home to Southern Styria and stay there for longer than a few days.

Last but not least I thank my girlfriend Doris for all the understanding every time I went to university earlier than planned just to come back later than expected.

Notation

Different notations in the mathematics and the physics community often leads to headache if a publication from the 'other' community is read. The notation used in this thesis is a compromise and tries to fit into both worlds.

Symbol	Meaning
\mathbb{R}	set of real numbers
$\ \cdot\ $	norm of the argument
$ \cdot $	modulus of the argument
$\mathcal{O}(\cdot)$	Landau symbol
Ω	unit sphere
θ	colatitude (polar angle)
φ	longitude (azimuth)
$d\Omega$	unit sphere surface element, $\sin\theta d\theta d\varphi$
e_θ, e_φ	unit vectors w.r.t. θ and φ
∇	Nabla operator
$\nabla \cdot u = \text{div}(u)$	divergence of u
Δ	Laplace operator
Δ^*	Beltrami operator
\mathbf{A}, \mathbf{B}	matrices (capital letters, bold)
\mathbf{v}, \mathbf{w}	vectors (lower case)
P_l	Legendre polynomial of degree l
P_l^m	associated Legendre function of major index l and minor index m
$Y_{l,m}$	real valued spherical harmonic of major index l and minor index m
$X_{l,m}$	spherical harmonics expansion coefficient to $Y_{l,m}$ of the quantity X
$\mathcal{F}_\Omega\{f\}$	spherical Fourier transform
$\delta_{i,j}$	Kronecker delta
$\delta(\cdot)$	delta distribution
\mathcal{T}	triangulation of the simulation domain
\mathcal{B}	dual box grid of \mathcal{T}
B	a box from the dual box grid \mathcal{B}
\mathcal{P}_s	space of (piecewise) polynomials of degree s
$\text{Bern}(x)$	Bernoulli function
N	number of degrees of freedom in $(\mathbf{x}, \varepsilon)$ -space
L	spherical harmonics expansion order

Symbol	Meaning
\hbar	Planck constant divided by 2π
ϵ	permittivity (not to be confused with energy ε)
ρ	space charge
$ q $	elementary charge
m^*	effective mass
ε	energy (not to be confused with permittivity ϵ)
\mathbf{x}	spatial coordinate within the device
\mathbf{k}	wave vector
\mathbf{p}	momentum vector
$\mathbf{v}(\mathbf{k})$	(group) velocity
\mathbf{F}	force
\mathbf{E}	electric field
$Q\{\cdot\}$	scattering operator
$Q^{\text{in}}\{\cdot\}$	in-scattering operator
$Q^{\text{out}}\{\cdot\}$	out-scattering operator
$Q^{\text{VR}}\{\cdot\}$	velocity-randomising scattering operator
η	index of the scattering process
$s(\mathbf{x}, \mathbf{k}, \mathbf{k}')$	scattering rate
$n(\mathbf{x}, t)$	particle density
$g(\varepsilon, \theta, \varphi)$	generalised density of states
$\mathbf{j}(\mathbf{x}, t)$	current density
$\langle \cdot \rangle$	moment with respect to the argument
$f(\mathbf{x}, \mathbf{k}, t)$	distribution function
$\tilde{f}(\mathbf{x}, \mathbf{k}, t)$	generalised energy distribution function
$H(f, \mathbf{x}, \mathbf{k})$	entropy
$H_l(f, \mathbf{x}, \mathbf{k})$	entropy factor
$\mathbf{v}_{l,m}(\varepsilon)$	projection of the (group) velocity
$\mathbf{\Gamma}_{l,m}(\varepsilon)$	spherical harmonics coupling term
V_T	thermal voltage

Introduction

Semiconductor devices have experienced a tremendous progress in the past decades, following Moore's Law which states that the number of components on a chip doubles approximately every two years. Since the chip sizes are held more or less constant, Moore's Law also states that feature sizes shrink exponentially. At present, MOSFETs with a channel length of only 20nm are in production.

Due to the high level of miniaturisation, computer aided design of semiconductor devices is widely used to optimise device layouts. While in the early years of the semiconductor industry macroscopic models such as the *drift diffusion model* or the *hydrodynamic model* have been sufficient for device simulation, this is not the case anymore for the small feature sizes used today. As long as quantum mechanical effects can be neglected, the microscopic behaviour of electrons is in accordance to the Boltzmann Transport Equation (BTE), which is considered to be the best classical description of electrons.

A direct solution of the BTE has been pursued for several decades and many ingenious techniques have been developed for this purpose. The most commonly used is the Monte Carlo method, primarily because the method is very flexible and allows details such as complicated band structures to be easily incorporated. The main disadvantage of the Monte Carlo method is its computational expense, especially when attempting to reduce the statistical noise in the low density tails of the distribution function.

As an alternative to the Monte Carlo method, the spherical harmonics expansion (SHE) method is considered in this thesis. In contrast to the former, the latter yields deterministic solutions of the BTE and is computationally less expensive. The spherical harmonics expansion method for the simulation of one-dimensional semiconductors was pioneered in the early 1990's [15, 17] and since then several scientists around the world have improved the method in several ways. We mention extensions to higher expansion orders [20, 21], higher spatial domains [16, 49] and more recently a stabilisation scheme [24]. The spherical harmonics expansion has also been encountered by mathematicians who have applied several limits to the BTE and the spherical harmonics expansion in order to derive simpler models [1, 4, 10, 12].

This thesis is organised as follows: In the first chapter the mathematical background of spherical harmonics is presented. Most results are known for more than a century, except for the modern approach of the spherical Fourier transform given in Section 1.5. The proof of Thm. 13 is contributed by the author of this thesis, mainly because it was not given in the referenced text-book.

In Chapter 2 the BTE is derived and the individual terms such as the scattering operator introduced. In contrast to the first chapter, where the emphasis is the introduction of a solid mathematical foundation, the second chapter aims at the introduction of the necessary topics from physics. There are no additional contributions by the author in this chapter.

The BTE is projected onto spherical harmonics in Chapter 3. Additionally, several intrinsic properties of the BTE are shown to be conserved after the projection onto spherical harmonics. Finally, the concept of entropies for the continuous SHE equations is introduced, which later serves as a stabilisation scheme for the discretised equations. The projection onto spherical harmonics follows the calculations from Jungemann et. al. [24], where several details of the calculation are filled in. Section 3.3 is compiled from several publications by Ringhofer [34–37].

Chapter 4 deals with the discretisation of the spherical harmonics expansion equations. At first a short introduction to the finite volume method is given, then the Scharfetter-Gummel scheme is briefly discussed to show the necessity of stabilisation schemes for semiconductor device modelling. The full Galerkin scheme for the discretisation of the spherical harmonics expansion equation in the simulation domain as well as the analytical formulae for simulations in two dimensions are contributed by the author of this thesis.

Based on the Galerkin scheme introduced in Chapter 4, complexity analysis is given in Chapter 5, which is the author's own work. On the one hand the coupling terms for coefficients of spherical harmonics of different order are shown to be sparse for spherical energy bands, while on the other hand it is shown that the memory requirements can be reduced considerably if the resulting system matrix is stored in a factored form. Matrix-vector multiplication is still possible with similar computational costs, which enables the use of iterative solvers for the solution of the system of linear equations.

In Chapter 6 some numerical results are given. It turned out that the SHE equations are much harder to implement than for example the drift-diffusion model, from which the initial guess for the potential is obtained. Moreover, preconditioners are observed to be necessary to keep the condition number of the resulting system sufficiently low.

Chapter 7 gives an outlook on further ideas for future work on this topic. This thesis then closes with a conclusion.

Chapter 1

Spherical Harmonics

Many physical quantities such as the potential induced by a charged particle show a radial symmetry, so that it is often convenient to use a spherical coordinate system. Thus, the Laplace equation in spherical coordinates was investigated in the 18th century, which has lead to so-called *spherical harmonics*. Conversely, spherical harmonics can be used for the approximation of functions on the sphere, thus they can be seen as three-dimensional extensions of sines and cosines used in Fourier analysis.

This chapter is devoted to the mathematical background of spherical harmonics and collects all important mathematical results for later use on the numerical solution of Boltzmann's transport equation using an expansion method. Most of the derivations and results have been known for more than a century, except for the last section, where newer results are collected. The reader is referred several times throughout this chapter to the literature for some of the more technical or lengthy proofs.

1.1 The Laplace Equation in Polar Coordinates

Gauss' Law states that the electric flux density \mathbf{D} is related to the space charge density ρ by

$$\nabla \cdot \mathbf{D} = \rho \quad (1.1)$$

Using the material equation $\mathbf{D} = \epsilon \mathbf{E}$, where ϵ is the permittivity and \mathbf{E} is the electric field, one finds with $\mathbf{E} = -\nabla\psi$ the governing equation for the electric potential φ :

$$-\nabla \cdot (\epsilon \nabla \psi) = \rho , \quad (1.2)$$

which is commonly termed *Poisson equation*. If the permittivity ϵ is a scalar, the homogeneous form is known as *Laplace's equation*

$$\Delta \psi = 0 . \quad (1.3)$$

In the following we write u instead of ψ , following mathematical conventions. The explicit form of the Laplace operator $\Delta := \nabla \cdot \nabla$ depends on the underlying coordinate system, in Cartesian coordinates (x, y, z) the operator is given as

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} . \quad (1.4)$$

As mentioned at the beginning of this chapter, for a large class of problems it is of advantage to use *spherical coordinates* (r, φ, θ) of the form

$$x = r \sin \theta \cos \varphi , \quad (1.5)$$

$$y = r \sin \theta \sin \varphi , \quad (1.6)$$

$$z = r \cos \theta \quad (1.7)$$

with $r \in [0, \infty)$, $\theta \in [0, \pi)$ and $\varphi \in [0, 2\pi)$. In spherical coordinates, Laplace's equation then reads

$$\Delta u = \frac{1}{r^2 \sin \theta} \left\{ \frac{\partial}{\partial r} \left(r^2 \sin \theta \frac{\partial u}{\partial r} \right) + \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial u}{\partial \theta} \right) + \frac{\partial}{\partial \varphi} \left(\frac{1}{\sin \theta} \frac{\partial u}{\partial \varphi} \right) \right\} = 0 . \quad (1.8)$$

We split the Laplacian into radial and angular parts by

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \Delta^* , \quad (1.9)$$

where the *Beltrami operator* Δ^* is given as

$$\Delta^* := \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} . \quad (1.10)$$

Note that for fixed r , e.g. $r = 1$, the Beltrami operator is the trace of the Laplace operator on the surface of a sphere with radius r .

An ansatz $u(r, \theta, \varphi) = R(r)\Theta(\theta)\Phi(\varphi)$ leads after substitution into (1.8) and division by $R(r)\Theta(\theta)\Phi(\varphi)$ to the equation

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \frac{1}{\Phi \sin^2 \theta} \frac{d^2 \Phi}{d\varphi^2} = 0 . \quad (1.11)$$

The first term is the only one which depends on r , thus it must be a constant, say k :

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) = k . \quad (1.12)$$

The solution of this equation can be found as

$$R(r) = Ar^l + Br^{-l-1} , \quad (1.13)$$

where A and B are arbitrary constants and $k = l(l+1)$. With this we can rewrite (1.11) as

$$l(l+1) \sin^2 \theta + \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \frac{1}{\Phi} \frac{d^2 \Phi}{d\varphi^2} = 0 . \quad (1.14)$$

This time the last term on the left hand side is the only that shows a dependence on φ , thus

$$\frac{1}{\Phi} \frac{d^2 \Phi}{d\varphi^2} = -m^2 . \quad (1.15)$$

for a (possibly complex valued) constant m . The general solution of Φ for real valued m is thus

$$\Phi(\varphi) = C \cos(m\varphi) + D \sin(m\varphi) . \quad (1.16)$$

Note that a right hand side m^2 instead of $-m^2$ in (1.15) would lead to hyperbolic functions in (1.16), so that $\Phi(\varphi)$ were globally bounded only in the case $C = D = 0$.

With the general solution for $\Phi(\varphi)$, (1.14) becomes

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \left(n(n+1) - \frac{m^2}{\sin^2 \theta} \right) = 0 , \quad (1.17)$$

which can be recast by setting $\mu = \cos \theta$, $\Theta = u$ into the form

$$\frac{d}{d\mu} \left[(1 - \mu^2) \frac{du}{d\mu} \right] + \left[l(l+1) - \frac{m^2}{1 - \mu^2} \right] u = 0 . \quad (1.18)$$

Provided that for a particular (entirely unrestricted) choice of m and l a solution u_l^m of (1.18) is found, the normal forms satisfying Laplace's equation are then given as

$$\begin{aligned} r^l u_l^m \cos(m\varphi), & \quad r^l u_l^m \sin(m\varphi), \\ r^{-l-1} u_l^m \cos(m\varphi), & \quad r^{-l-1} u_l^m \sin(m\varphi). \end{aligned}$$

In the following we will confine ourselves to the case in which m and l are positive integers.

1.2 Legendre Polynomials

In this section we consider the case $m = 0$ and nonnegative integers l in (1.18), thus

$$\frac{d}{d\mu} \left[(1 - \mu^2) \frac{du}{d\mu} \right] + l(l+1)u = 0 , \quad (1.19)$$

which is known as *Legendre's equation*. This equation is often written equivalently in the form

$$(1 - \mu^2) \frac{d^2 u}{d\mu^2} - 2\mu \frac{du}{d\mu} + l(l+1)u = 0 . \quad (1.20)$$

By an ansatz $u = a_0 + a_1\mu + a_2\mu^2 + \dots$ and an ansatz $u = \mu^m + \alpha_2\mu^{m-2} + \alpha_4\mu^{m-4} + \dots$ the following theorem can be shown [22]:

Theorem 1. *The complete solution of Legendre's equation with positive integers l is*

$$u = AP_l(\mu) + BQ_l(\mu) , \quad (1.21)$$

where A and B denote arbitrary constants. The expression $P_l(\mu)$ is an algebraic function of μ , of degree l , and is given by

$$P_l(\mu) = \frac{(2l)!}{2^l l! l!} \mu^l F \left(-\frac{l}{2}, \frac{1-l}{2}; \frac{1}{2} - l; \frac{1}{\mu^2} \right) , \quad (1.22)$$

where $F(\cdot, \cdot; \cdot; \cdot)$ denotes the hypergeometrical series. The second solution $Q_l(\mu)$ is singular at $|\mu| = 1$

$P_l(\mu)$ is called *Legendre polynomial of degree l* , Q_l is called *Legendre function of the second kind*; explicit formulae for Q_l for the cases $|\mu| < 1$ and $|\mu| > 1$ can be given [22, 44].

From (1.22) it is hard to see that $P_l(\mu)$ is indeed a polynomial, while it can be readily seen from the following result:

Theorem 2 (Rodrigues' Theorem). *For Legendre polynomials $P_l(\mu)$ with $l = \{0, 1, 2, \dots\}$ as given in (1.22) there holds*

$$P_l(\mu) = \frac{1}{2^l l!} \frac{d^l}{d\mu^l} (\mu^2 - 1)^l \quad \mu \in [-1, 1] . \quad (1.23)$$

Proof. We refer to the literature [22, 44] for a proof of Rodrigues' Theorem. \square

Rodrigues' Theorem is an important analytical result which allows to show many other important properties of Legendre polynomials, such as orthogonality:

Theorem 3 (Orthogonality of Legendre Polynomials). *For positive integer l there holds*

$$\int_{-1}^1 \mu^k P_l(\mu) d\mu = 0 , \quad \forall k = 0, 1, \dots, l-1 . \quad (1.24)$$

Consequently, Legendre polynomials are orthogonal with respect to the inner product of $L^2([-1, 1])$:

$$\int_{-1}^1 P_k(\mu) P_l(\mu) d\mu = \begin{cases} 0, & k \neq l, \\ 2/(2k+1), & k = l, \end{cases} \quad \forall k, l \in \{0, 1, 2, \dots\} . \quad (1.25)$$

Proof. We substitute Rodrigues' expression for $P_l(\mu)$ into (1.24) and obtain for $0 \leq k < l$ using integration by parts and the fact that all derivatives of $(\mu^2 - 1)^l$ of lower degree than l vanish at $\mu = \pm 1$, that

$$\begin{aligned} \int_{-1}^1 \mu^k P_l(\mu) d\mu &= \frac{1}{2^l l!} \int_{-1}^1 \mu^k \frac{d^l}{d\mu^l} (\mu^2 - 1)^l d\mu \\ &= \frac{1}{2^l l!} \left[\mu^k \frac{d^{l-1}}{d\mu^{l-1}} (\mu^2 - 1)^l \right]_{-1}^1 - \frac{k}{2^l l!} \int_{-1}^1 \mu^{k-1} \frac{d^{l-1}}{d\mu^{l-1}} (\mu^2 - 1)^l d\mu \\ &= \frac{(-1)^k k}{2^l l!} \int_{-1}^1 \mu^{k-1} \frac{d^{l-1}}{d\mu^{l-1}} (\mu^2 - 1)^l d\mu \\ &\quad \vdots \\ &= \frac{(-1)^k k!}{2^l l!} \int_{-1}^1 \frac{d^{l-k}}{d\mu^{l-k}} (\mu^2 - 1)^l d\mu \\ &= \frac{(-1)^k k!}{2^l l!} \left[\frac{d^{l-k-1}}{d\mu^{l-k-1}} (\mu^2 - 1)^l \right]_{-1}^1 \\ &= 0 . \end{aligned}$$

This shows (1.24). The orthogonality of Legendre polynomials follows immediately from the linearity of the integral, since $P_k(\mu)$ is the sum of a number of terms of monomials

in μ with order smaller or equal to k each. Finally, we use again Rodrigues' formula and integration by parts to find

$$\begin{aligned} \int_{-1}^1 [P_l(\mu)]^2 d\mu &= \frac{1}{2^{2l} l! l!} \int_{-1}^1 \frac{d^l}{d\mu^l} (\mu^2 - 1)^l \frac{d^l}{d\mu^l} (\mu^2 - 1)^l d\mu \\ &= \frac{(-1)^l}{2^{2l} l! l!} \int_{-1}^1 (\mu^2 - 1)^l \frac{d^{2l}}{d\mu^{2l}} (\mu^2 - 1)^l d\mu \\ &= \frac{(2l)!}{2^{2l} l! l!} \int_{-1}^1 (1 - \mu^2)^l d\mu. \end{aligned}$$

Setting $t = (1 + \mu)/2$, we obtain

$$\begin{aligned} \int_{-1}^1 [P_l(\mu)]^2 d\mu &= 2 \frac{(2l)!}{l! l!} \int_0^1 t^l (1 - t)^l dt \\ &= 2 \frac{(2l)!}{l! l!} \mathcal{B}(l + 1, l + 1) = \frac{2}{2l + 1}, \end{aligned}$$

by the use of the beta function $\mathcal{B}(\cdot, \cdot)$. □

Even though Rodrigues' formula is a valuable analytical tool, it is not suited for numerical evaluations. However, since Legendre polynomials are a set of orthogonal polynomials with respect to the inner product of the Hilbert space $L^2([-1, 1])$, there exists a three-term recurrence relation [14], which is highly attractive for efficient numerical evaluations:

Theorem 4. *The following relation between three consecutive Legendre polynomials holds:*

(i)

$$lP_l(\mu) - (2l - 1)\mu P_{l-1}(\mu) + (l - 1)P_{l-2} = 0. \quad (1.26)$$

(ii)

$$\frac{dP_l(\mu)}{d\mu} - \frac{dP_{l-2}}{d\mu} = (2l - 1)P_{l-1}(\mu). \quad (1.27)$$

Proof. The first recursion is proved in a more general context in Thm. 6. The second recursion is obtained either by the use of Rodrigues' Theorem or by clever manipulation of Legendre's equation [14, 22]. □

We note that $P_0(\mu) = 1$, $P_1(\mu) = \mu$, so all Legendre polynomials of higher order can now be obtained with (1.26).

1.3 Associated Legendre Functions

In Section 1.1 we have found from separation of variables that Laplace's equation is satisfied by functions of the form

$$\begin{aligned} r^l u_l^m \cos(m\varphi), & \quad r^l u_l^m \sin(m\varphi), \\ r^{-l-1} u_l^m \cos(m\varphi), & \quad r^{-l-1} u_l^m \sin(m\varphi), \end{aligned}$$

where u_l^m satisfies

$$\frac{d}{d\mu} \left[(1 - \mu^2) \frac{du}{d\mu} \right] + \left[l(l+1) - \frac{m^2}{1 - \mu^2} \right] u = 0 . \quad (1.28)$$

While Legendre polynomials are obtained as solutions in the case $m = 0$, we will now consider the case that both l and m are nonnegative integers with $l \geq m$.

Setting $u = v(\mu^2 - 1)^{m/2}$, a brief calculation shows that v satisfies the equation

$$(1 - \mu^2) \frac{d^2 v}{d\mu^2} - 2(m+1)\mu \frac{dv}{d\mu} + (l-m)(l+m+1)v = 0 . \quad (1.29)$$

On the other hand, if we differentiate Legendre's equation (1.19) m times, we find that

$$(1 - \mu^2) \frac{d^{m+2} u}{d\mu^{m+2}} - 2(m+1)\mu \frac{d^{m+1} u}{d\mu^{m+1}} + (l-m)(n+m+1) \frac{d^m u}{d\mu^m} = 0 , \quad (1.30)$$

thus $d^m u / d\mu^m$ satisfies (1.29) and we have

$$v = A \frac{d^m P_l}{d\mu^m} + B \frac{d^m Q_l}{d\mu^m} , \quad (1.31)$$

where A and B are free parameters. The complete solution of (1.28) is therefore

$$u = A(\mu^2 - 1)^{m/2} \frac{d^m P_l}{d\mu^m} + B(\mu^2 - 1)^{m/2} \frac{d^m Q_l}{d\mu^m} , \quad (1.32)$$

which motivates the following definition for real valued arguments $-1 \leq \mu \leq 1$

Definition 1. We denote with

$$P_l^m(\mu) := (-1)^m (1 - \mu^2)^{m/2} \frac{d^m P_l}{d\mu^m} = \frac{(-1)^m}{2^l l!} (1 - \mu^2)^{m/2} \frac{d^{l+m}}{d\mu^{l+m}} (\mu^2 - 1)^l \quad (1.33)$$

the associated Legendre function of first kind and with

$$Q_l^m(\mu) := (-1)^m (1 - \mu^2)^{m/2} \frac{d^m Q_l}{d\mu^m} \quad (1.34)$$

the associated Legendre function of second kind. We call the lower index l major index and the upper index m minor index.

The term $(-1)^m$ results from extensions to the complex plane and is – motivated from quantum physics – called *Condon-Shortley phase*.

Definition (1.33) makes sense for integers $m \leq l$ only. So far we have considered the case $m \geq 0$ only, but changing the sign of m does not alter (1.28). Thus, with similar arguments for $(\mu^2 - 1)^{-m/2} v$, we find the most general form of v as

$$\frac{d^m v}{d\mu^m} = A P_l(\mu) + B Q_l(\mu) . \quad (1.35)$$

Consequently, we define for $m > 0$

$$P_l^{-m} := (1 - \mu^2)^{-m/2} \int_1^\mu \cdots \int_1^\mu P_l(\mu) d\mu \cdots d\mu \quad (1.36)$$

$$\stackrel{m \leq l}{=} \frac{(-1)^{-m}}{2^l l!} (1 - \mu^2)^{-m/2} \frac{d^{l-m}}{d\mu^{l-m}} (\mu^2 - 1)^l, \quad (1.37)$$

where the integrals are m -fold. The second equality has to be taken with a grain of salt: While the integral expression on the left hand side can be extended to the case $m > l$ directly, the notation used on the right hand side does not. Nevertheless, we observe that (1.33) is thus valid for all integers m with $|m| \leq l$.

From the explicit form (1.33) it is not clear whether or how $P_l^m(\mu)$ is related to $P_l^{-m}(\mu)$, $0 \leq m \leq l$. We can observe in a first step that for $m > 0$ both $(1 - \mu^2)^{m/2} P_l^m(\mu)$ and $(1 - \mu^2)^{m/2} P_l^{-m}(\mu)$ are polynomials of degree $l + m$, but we can say more:

Theorem 5. *Let $0 \leq m \leq l$ with nonnegative integers m and l . Then*

$$P_l^{-m}(\mu) = (-1)^m \frac{(l - m)!}{(l + m)!} P_l^m(\mu). \quad (1.38)$$

Proof. The proportionality constant can be checked easily for the coefficients of μ^{l+m} after multiplication of (1.38) with $(1 - \mu^2)^{m/2}$. For a proof of the proportionality of the remaining coefficients, the reader is referred to the textbook of Hobson [22]. \square

As for Legendre polynomials, numerical evaluation of associated Legendre functions is carried out by the use of recursion formulae:

Theorem 6. *For three consecutive Legendre polynomials the recurrence relation*

$$(l - m + 2)P_{l+2}^m(\mu) = (2l + 3)\mu P_{l+1}^m(\mu) - (l + m + 1)P_l^m(\mu), \quad 0 \leq m \leq l \quad (1.39)$$

holds. The recursion terminates with the explicit expressions

$$P_m^m(\mu) = (-1)^m \frac{(2m)!}{2^m m!} (1 - \mu^2)^{m/2}, \quad (1.40)$$

$$P_{m+1}^m(\mu) = (2m + 1)\mu P_m^m(\mu). \quad (1.41)$$

Proof. The expression for $P_m^m(\mu)$ follows immediately from the $2m$ -th derivative of $(\mu^2 - 1)^m$. Similarly, we find after differentiation

$$P_{m+1}^m(\mu) = (-1)^m \frac{(2m + 2)!}{2^{m+1}(m + 1)!} \mu (1 - \mu^2)^{m/2} = (2m + 1)\mu P_m^m(\mu).$$

To prove the recurrence relation, we differentiate (1.26) m times to obtain

$$l \frac{d^m P_l(\mu)}{d\mu^m} - (2l - 1)\mu \frac{d^m P_{l-1}(\mu)}{d\mu^m} - (2l - 1)m \frac{d^{m-1} P_{l-1}(\mu)}{d\mu^{m-1}} + (l - 1) \frac{d^m P_{l-2}(\mu)}{d\mu^m} = 0.$$

Differentiating (1.27) $m - 1$ times, we find

$$\frac{d^m P_l(\mu)}{d\mu^m} - \frac{d^m P_{l-2}(\mu)}{d\mu^m} = (2l - 1) \frac{d^{m-1} P_{l-1}(\mu)}{d\mu^{m-1}},$$

so that elimination of $d^{m-1}P_{l-1}(\mu)/d\mu^{m-1}$ yields

$$(l-m)\frac{d^m P_l(\mu)}{d\mu^m} - (2l-1)\mu\frac{d^m P_{l-1}(\mu)}{d\mu^m} + (l+m-1)\frac{d^m P_{l-2}(\mu)}{d\mu^m} = 0 .$$

Multiplication with $(-1)^m(1-\mu^2)^{m/2}$ together with a shift of indices gives (1.39). \square

We have already seen in Thm. 3 that the set $\{P_k^0\}_{k=0}^l$ (recall that for the case $m=0$ the associated Legendre functions coincide with the Legendre polynomials) is orthogonal with respect to $L^2([-1, 1])$. In fact, this is true for all $0 \leq m \leq l$:

Theorem 7 (Orthogonality of associated Legendre functions). *For fixed m, l , with $0 \leq m \leq l$, and integers $0 \leq r, s \leq n$, the associated Legendre functions are orthogonal with respect to the inner product on $L^2([-1, 1])$:*

$$\int_{-1}^1 P_r^m(\mu) P_s^m(\mu) d\mu = \begin{cases} 0, & r \neq s, \\ \frac{(r+m)!}{(r-m)!} \frac{2}{2r+1}, & r = s, \end{cases} \quad \forall r, s \in \{0, 1, 2, \dots\} . \quad (1.42)$$

Proof. All $P_s^m(\mu)$ with $0 \leq m \leq s$ fulfil (1.28), thus:

$$\frac{d}{d\mu} \left[(1-\mu^2) \frac{dP_s^m(\mu)}{d\mu} \right] + \left[s(s+1) - \frac{m^2}{1-\mu^2} \right] P_s^m(\mu) = 0 . \quad (1.43)$$

Multiplication with $P_r^m(\mu)$ and integration over $[-1, 1]$ leads to

$$\int_{-1}^1 P_r^m(\mu) \frac{d}{d\mu} \left[(1-\mu^2) \frac{dP_s^m(\mu)}{d\mu} \right] + P_r^m(\mu) \left[s(s+1) - \frac{m^2}{1-\mu^2} \right] P_s^m(\mu) d\mu = 0 . \quad (1.44)$$

Integration by parts of the first summand in the integrand yields

$$\int_{-1}^1 \frac{dP_r^m(\mu)}{d\mu} (\mu^2 - 1) \frac{dP_s^m(\mu)}{d\mu} + P_r^m(\mu) \left[s(s+1) - \frac{m^2}{1-\mu^2} \right] P_s^m(\mu) d\mu = 0 . \quad (1.45)$$

Interchanging r and s we find

$$\int_{-1}^1 \frac{dP_r^m(\mu)}{d\mu} (\mu^2 - 1) \frac{dP_s^m(\mu)}{d\mu} + P_r^m(\mu) \left[r(r+1) - \frac{m^2}{1-\mu^2} \right] P_s^m(\mu) d\mu = 0 , \quad (1.46)$$

so that a subtraction of the two equations yields

$$[s(r+1) - r(r+1)] \int_{-1}^1 P_r^m(\mu) P_s^m(\mu) d\mu = 0 , \quad (1.47)$$

and we immediately deduce

$$\int_{-1}^1 P_r^m(\mu) P_s^m(\mu) d\mu = 0 , \quad r \neq s . \quad (1.48)$$

To find the value of the integral in case $r = s$, we first replace m by $m-1$ as well as u by the Legendre polynomial $P_s(\mu)$ in (1.30):

$$(1-\mu^2) \frac{d^{m+1} P_s(\mu)}{d\mu^{m+1}} - 2m\mu \frac{d^m P_s(\mu)}{d\mu^m} + (l-m+1)(l+m) \frac{d^{m-1} P_s(\mu)}{d\mu^{m-1}} = 0 . \quad (1.49)$$

Multiplication with $(1 - \mu^2)^{m-1}$ leads to

$$(1 - \mu^2)^m \frac{d^{m+1}P_s(\mu)}{d\mu^{m+1}} - 2m\mu(1 - \mu^2)^{m-1} \frac{d^m P_s(\mu)}{d\mu^m} + (l - m + 1)(l + m)(1 - \mu^2)^{m-1} \frac{d^{m-1}P_s(\mu)}{d\mu^{m-1}} = 0, \quad (1.50)$$

which can be rewritten in the form

$$\frac{d}{d\mu} \left[(1 - \mu^2)^m \frac{d^m P_s(\mu)}{d\mu^m} \right] = -(l - m + 1)(l + m)(1 - \mu^2)^{m-1} \frac{d^{m-1}P_s(\mu)}{d\mu^{m-1}}. \quad (1.51)$$

With this we can manipulate using integration per parts

$$\begin{aligned} \int_{-1}^1 P_r^m(\mu) P_r^m(\mu) d\mu &= \int_{-1}^1 \frac{d^m P_r(\mu)}{d\mu^m} (1 - \mu)^m \frac{d^m P_r(\mu)}{d\mu^m} d\mu \\ &= - \int_{-1}^1 \frac{d^{m-1} P_r(\mu)}{d\mu^{m-1}} \frac{d}{d\mu} \left[(1 - \mu)^m \frac{d^m P_r(\mu)}{d\mu^m} \right] d\mu \\ &= (r - m + 1)(r + m) \int_{-1}^1 \frac{d^{m-1} P_r(\mu)}{d\mu^{m-1}} (1 - \mu)^{m-1} \frac{d^{m-1} P_r(\mu)}{d\mu^{m-1}} d\mu \\ &\quad \vdots \\ &= \frac{(r + m)!}{(r - m)!} \int_{-1}^1 P_r(\mu) P_r(\mu) d\mu \\ &= \frac{(r + m)!}{(r - m)!} \frac{2}{2r + 1}. \end{aligned}$$

□

We note that by the use of the results from Thm. 5, Thm. 7 can be shown to hold unaltered in the more general case $-l \leq m \leq l$ [3].

Instead of keeping the minor index m of $P_l^m(\mu)$ fixed and derive recursions with respect to l , recursions with varying minor index m can also be derived and will be used in Chapter 5:

Lemma 1. *The following recurrence relations for associated Legendre functions with $0 \leq m \leq l$ hold:*

(i)

$$\mu P_l^m(\mu) = (l - m + 1)(1 - \mu^2)^{1/2} P_l^{m-1}(\mu) + P_{l-1}^m(\mu) \quad (1.52)$$

(ii)

$$P_{l+1}^m(\mu) = \mu P_l^m(\mu) + (l + m)(1 - \mu^2)^{1/2} P_l^{m-1}(\mu) \quad (1.53)$$

(iii)

$$(l + m + 1)\mu P_l^m(\mu) = (l - m + 1)P_{l+1}^m(\mu) + (1 - \mu^2)^{1/2} P_l^{m+1}(\mu) \quad (1.54)$$

(iv)

$$(l+m)P_{l-1}^m(\mu) = (1-\mu^2)^{1/2}P_l^{m+1}(\mu) + (l-m)\mu P_l^m(\mu) \quad (1.55)$$

(v)

$$(1-\mu^2)\frac{dP_l^m(\mu)}{d\mu} = (l+m)P_{l-1}^m(\mu) - l\mu P_l^m(\mu), \quad (1.56)$$

In case that the minor index s is larger than the major index r , we set $P_r^s \equiv 0$, cf. (1.33).

Proof. These recurrence relations can be verified directly by substitution of the explicit representation (1.33) into (1.52) - (1.56) and application of suitable rules for differentiation. \square

1.4 Orthogonal Functions on the Sphere

Legendre polynomials and associated Legendre functions have been introduced to cover the angular dependence of harmonic functions on the colatitude (or polar angle) θ . The dependence of a harmonic function on the sphere on the longitude (or azimuth) φ is resolved by sines and cosines, so that in accordance with the ansatz $u(r, \theta, \varphi) = R(r)\Theta(\theta)\Phi(\varphi)$ we define

Definition 2 (Real valued spherical harmonics). *The real valued spherical harmonics are defined as the space induced by the set*

$$Y_{l,m}(\theta, \varphi) = \begin{cases} \sqrt{\frac{1}{\pi} \frac{(l-m)!}{(l+m)!} \frac{2l+1}{2}} P_l^m(\cos \theta) \cos(m\varphi), & m > 0, \\ \sqrt{\frac{1}{2\pi} \frac{2l+1}{2}} P_l^0(\cos \theta), & m = 0, \\ \sqrt{\frac{1}{\pi} \frac{(l+m)!}{(l-m)!} \frac{2l+1}{2}} P_l^{-m}(\cos \theta) \sin(m\varphi), & m < 0, \end{cases} \quad (1.57)$$

for $l = 0, 1, \dots$ and $-l \leq m \leq l$.

Using the orthogonality results from the preceding sections, we obtain

Theorem 8 (Orthonormality of Spherical Harmonics). *The spherical harmonics as defined in (1.57) are orthonormal with respect to the inner product of $L^2(\Omega)$.*

Proof. Writing $N_{l,m}$ for the normalisation constants in (1.57) and putting

$$g_m(\varphi) = \begin{cases} \cos(m\varphi), & m > 0, \\ 1, & m = 0, \\ \sin(m\varphi), & m < 0, \end{cases} \quad (1.58)$$

we can directly show

$$\begin{aligned} \int_0^\pi \int_0^{2\pi} Y_{l,m}(\theta, \varphi) Y_{l',m'}(\theta, \varphi) \sin \theta \, d\varphi \, d\theta &= \\ &= N_{l,m} N_{l',m'} \int_0^\pi P_l^{|m|}(\cos \theta) P_{l'}^{|m'|}(\cos \theta) \sin \theta \, d\theta \times \int_0^{2\pi} g_m(\varphi) g_{m'}(\varphi) \, d\varphi \\ &= N_{l,m} N_{l',m'} \int_{-1}^1 P_l^{|m|}(\mu) P_{l'}^{|m'|}(\mu) \, d\mu \times \int_0^{2\pi} g_m(\varphi) g_{m'}(\varphi) \, d\varphi \\ &= \delta_{l,l'} \delta_{m,m'}, \end{aligned}$$

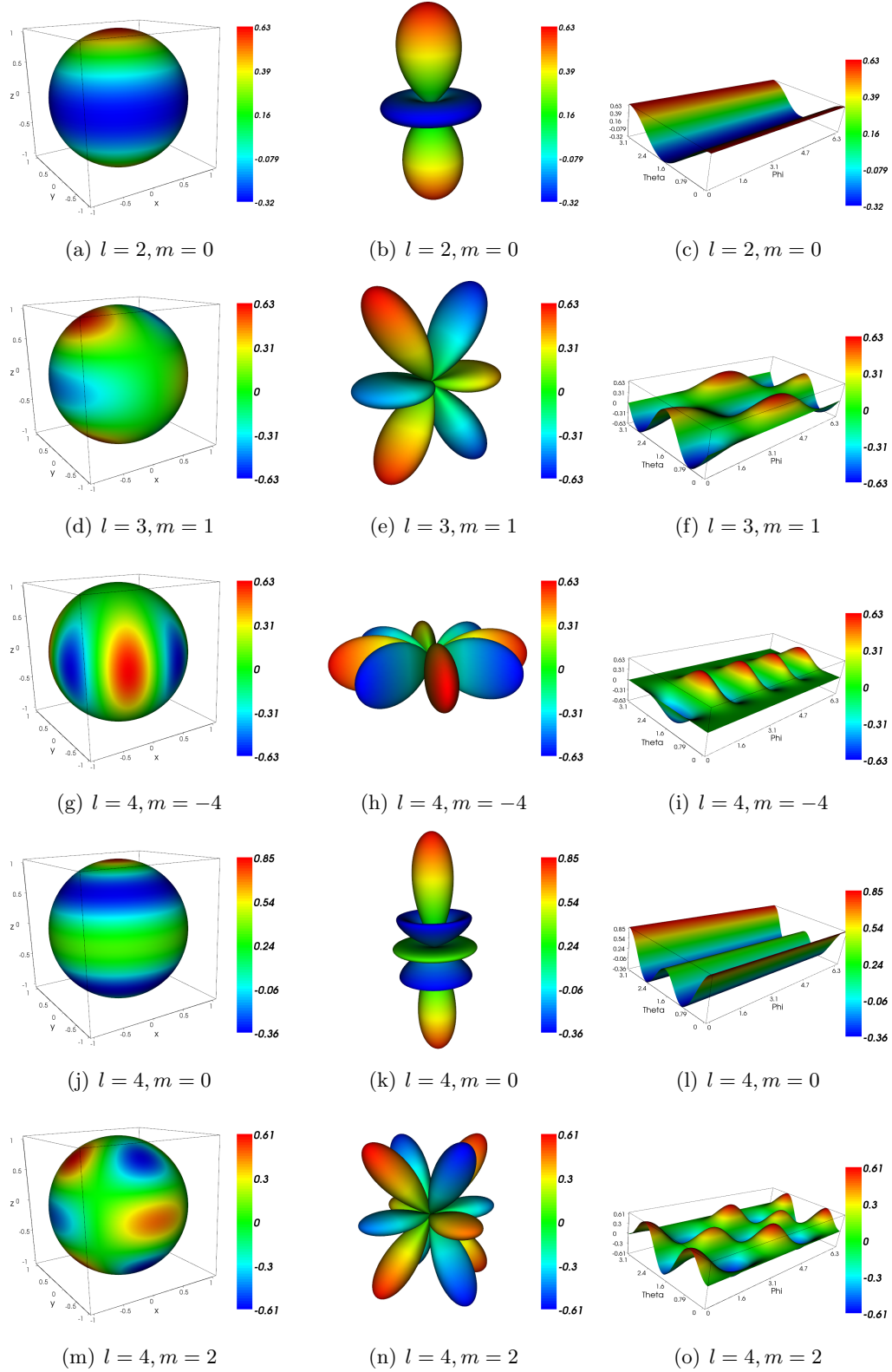


Figure 1.1: Selected real valued spherical harmonics $Y_{l,m}$ for small values of l and m . Evaluation is carried out over the sphere, in polar coordinates and as three-dimensional plot over the φ - θ -plane.

where $\delta_{k,k'}$ denotes the *Kronecker delta*, $\mu = \cos \theta$ and we have used the well known orthogonality of the set $\{g_m\}$ (cf. Fourier series). \square

It is also possible to define complex-valued spherical harmonics. In fact, the more natural way even is to allow complex valued values for the associated Legendre functions and define complex-valued spherical harmonics. The real-valued spherical harmonics can then be obtained as restriction of the complex plane to the real axis.

Typically, spherical harmonics are divided into three groups (cf. Fig. 1.1):

- Zonal harmonics: Spherical harmonics with $m = 0$
- Sectoral harmonics: Spherical harmonics with $|m| = l > 0$
- Tesseral harmonics: All spherical harmonics which are neither zonal nor sectorial.

If a spherical harmonic $Y_{l,m}$ is evaluated at a point \mathbf{p} with angular coordinates (θ, φ) , it is possible to deduce the value of the spherical harmonic at the opposing point \mathbf{p}' with coordinates $(\pi - \theta, \pi + \varphi)$:

Lemma 2 (Evaluation at the opposing point). *There holds*

$$Y_{l,m}(\theta, \varphi) = (-1)^l Y_{l,m}(\pi - \theta, \pi + \varphi) . \quad (1.59)$$

In other words: The parity of spherical harmonics with respect to points in (x, y, z) -coordinates is given by the parity of the major index l .

Proof. First we note that there holds

$$Y_{l,m}(\pi - \theta, \pi + \varphi) = N_{l,m} P_l^{|m|}(\cos(\pi - \theta)) g_m(\varphi) = N_{l,m} P_l^{|m|}(-\cos(\theta)) g_m(\varphi) , \quad (1.60)$$

where $N_{l,m}$ denotes the normalisation constant and $g_m(\varphi)$ is a shorthand notation as defined in (1.58). From the definition of the associated Legendre functions (1.33) it can easily be seen that the parity is even whenever $l - m$ is even, thus

$$Y_{l,m}(\pi - \theta, \pi + \varphi) = (-1)^{l-m} Y_{l,m}(\theta, \pi + \varphi) . \quad (1.61)$$

Since

$$\begin{aligned} \cos(m(\pi + \varphi)) &= \cos(m\pi + m\varphi) = (-1)^m \cos(m\varphi) \\ \sin(m(\pi + \varphi)) &= \sin(m\pi + m\varphi) = (-1)^m \sin(m\varphi) , \end{aligned}$$

we see that $g(\pi + \varphi) = (-1)^m g(\varphi)$ and thus obtain

$$Y_{l,m}(\pi - \theta, \pi + \varphi) = (-1)^{l-m} Y_{l,m}(\theta, \pi + \varphi) = (-1)^l Y_{l,m}(\theta, \varphi) . \quad (1.62)$$

\square

Similar results can be given for the partial derivatives:

Lemma 3. *There holds*

$$\frac{\partial Y_{l,m}(\theta, \varphi)}{\partial \theta} = (-1)^{l+1} \frac{\partial Y_{l,m}(\pi - \theta, \pi + \varphi)}{\partial \theta} , \quad (1.63)$$

$$\frac{\partial Y_{l,m}(\theta, \varphi)}{\partial \varphi} = (-1)^l \frac{\partial Y_{l,m}(\pi - \theta, \pi + \varphi)}{\partial \varphi} . \quad (1.64)$$

Proof. The second statement is clear since derivatives of trigonometric functions are again trigonometric functions and hence the same reasoning as for Lemma 2 applies.

To show (1.63), we have to consider the parity of the derivative of associated Legendre functions:

$$\begin{aligned} \frac{dP_l^m(\mu)}{d\mu} &= \frac{d}{d\mu} \left[\frac{(-1)^m}{2^l l!} (1 - \mu^2)^{m/2} \frac{d^{l+m}}{d\mu^{l+m}} (\mu^2 - 1)^l \right] \\ &= \frac{m(-1)^m}{2^{l+1} l!} \mu (1 - \mu^2)^{m/2-1} \frac{d^{l+m}}{d\mu^{l+m}} (\mu^2 - 1)^l \\ &\quad + \frac{(-1)^m}{2^l l!} (1 - \mu^2)^{m/2} \frac{d^{l+m+1}}{d\mu^{l+m+1}} (\mu^2 - 1)^l. \end{aligned}$$

The parity of the addends is even whenever $l + m$ is odd and vice versa, thus

$$\frac{\partial Y_{l,m}(\theta, \varphi)}{\partial \theta} = (-1)^{l-m+1} \frac{\partial Y_{l,m}(\pi - \theta, \varphi)}{\partial \theta}. \quad (1.65)$$

As in the proof of the previous Lemma, changing the second argument from φ to $\pi + \varphi$ leads to an additional factor $(-1)^m$, thus we arrive at (1.63). \square

Apart from orthogonality of spherical harmonics in $L^2(\Omega)$, we additionally obtain from the preceding sections the following

Theorem 9 (Eigenfunctions of the Beltrami operator). *Any spherical harmonic $Y_{l,m}$, $l = 0, 1, \dots$, $m = -l, \dots, l$ is an eigenfunction of the Beltrami operator Δ^* corresponding to the eigenvalue $-l(l+1)$.*

Proof. We substitute an harmonic function of the form $H(r, \theta, \varphi) = r^l Y_{l,m}(\theta, \varphi)$ into the Laplace equation and obtain using (1.9)

$$\begin{aligned} 0 = \Delta H &= r^{l-2} l(l+1) Y_{l,m}(\theta, \varphi) + r^{l-2} \Delta^* Y_{l,m}(\theta, \varphi), \\ \iff \Delta^* Y_{l,m}(\theta, \varphi) &= -l(l+1) Y_{l,m}(\theta, \varphi), \end{aligned}$$

hence $Y_{l,m}$ is an eigenfunction of the Beltrami operator Δ^* with eigenvalue $-l(l+1)$. \square

Several recursion formulas for spherical harmonics can directly be derived in a straightforward manner from those of associated Legendre functions. A more involved result is the following:

Theorem 10 (Addition Theorem). *For two points on the unit sphere with polar angles (θ, φ) and (θ', φ') , there holds for nonnegative l :*

$$P_l(\cos \alpha) = \frac{4\pi}{2l+1} \sum_{m=-l}^l Y_{l,m}(\theta, \varphi) Y_{l,m}(\theta', \varphi'), \quad (1.66)$$

where

$$\cos \alpha = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\varphi - \varphi') \quad (1.67)$$

is the angle between these two points.

Proof. For a proof we refer to the literature [22, 44]. \square

For later use we note that in the special case $(\theta, \varphi) = (\theta', \varphi')$, we have

$$1 = P_l(1) = \frac{4\pi}{2l+1} \sum_{m=-l}^l (Y_{l,m}(\theta, \varphi))^2, \quad (1.68)$$

where the first equality can be seen from the recurrence (1.26).

In the second half of the twentieth century a different approach to the (formal) construction of spherical harmonics became popular among mathematicians: Let $\text{Hom}_l(\mathbb{R}^3)$ denote the space of all homogeneous¹ polynomials of degree l . Next, we consider all harmonic homogeneous polynomials:

Definition 3. We denote

$$\text{Harm}_l(\mathbb{R}^k) = \left\{ p_l \in \text{Hom}_l(\mathbb{R}^3) : \Delta p_l(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathbb{R}^k \right\}, \quad (1.69)$$

where $k = 1, 2, 3, \dots$, the space of all polynomials of degree l that are harmonic and homogeneous.

Any $p_l \in \text{Harm}_l(\mathbb{R}^3)$ can be written in the form

$$p_l(\mathbf{x}) = p_l(x_1, x_2, x_3) = \sum_{j=0}^l q_{l-j}(x_1, x_2) x_3^j, \quad (1.70)$$

where $q_{l-j} \in \text{Hom}_{l-j}(\mathbb{R}^2)$. By using that all polynomials of degree zero or one are homogeneous, we find

$$0 = \Delta p_l(\mathbf{x}) = \sum_{j=0}^l \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) q_{l-j}(x_1, x_2) x_3^j \quad (1.71)$$

$$+ \sum_{j=0}^{l-2} q_{l-j-2}(x_1, x_2) (j+2)(j+1) x_3^j. \quad (1.72)$$

Hence by comparison of coefficients of x_3^j , we find the recursion relation

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) q_{l-j}(x_1, x_2) + (j+2)(j+1) q_{l-j-2}(x_1, x_2) = 0, \quad j = 0, \dots, l-2. \quad (1.73)$$

Thus, all q_{l-j} for $j = 2, \dots, n$ are determined for given q_l and q_{l-1} . This recursion formula actually allows us to find the dimension of $\text{Harm}_l(\mathbb{R}^3)$:

Lemma 4. The dimension of $\text{Harm}_l(\mathbb{R}^3)$ is $2l+1$.

Proof. As we have just seen, every homogeneous, harmonic polynomial $p(x)$ is determined by homogeneous $q_l \in \text{Hom}_l(\mathbb{R}^2)$ and $q_{l-1} \in \text{Hom}_{l-1}(\mathbb{R}^2)$, thus

$$\dim(\text{Harm}_l(\mathbb{R}^3)) = \dim(\text{Hom}_l(\mathbb{R}^2)) + \dim(\text{Hom}_{l-1}(\mathbb{R}^2)) \quad (1.74)$$

Since the dimension of $\text{Hom}_l(\mathbb{R}^2)$ and $\text{Hom}_{l-1}(\mathbb{R}^2)$ can easily be directly found as $l+1$ and l respectively, we immediately obtain the dimension of $\text{Harm}_l(\mathbb{R}^3)$ to be $2l+1$. \square

¹A polynomial p_l of degree l is called homogeneous if $p_l(\lambda \mathbf{x}) = \lambda^l p_l(\mathbf{x})$ for all $\lambda \in \mathbb{R}$ and all $\mathbf{x} \in \mathbb{R}^3$.

An alternative definition to Def. 2 consequently is

Definition 4 (Alternative definition of real valued spherical harmonics). *Spherical harmonics are obtained as a restriction of $\text{Harm}_l(\mathbb{R}^3)$ to the sphere Ω .*

With the use of the trigonometric identities

$$1 = \sin^2 \theta + \cos^2 \theta \quad (1.75)$$

$$\cos(m\varphi) = \sum_{k=0}^{[m/2]} \binom{m}{2k} (-1)^k \cos^{m-2k}(\varphi) \sin^{2k}(\varphi) \quad (1.76)$$

$$\sin(m\varphi) = \sum_{k=0}^{[m/2]} \binom{m}{2k+1} (-1)^k \cos^{m-2k-1}(\varphi) \sin^{2k+1}(\varphi), \quad (1.77)$$

where $[x]$ denotes the largest integer smaller or equal to x , in Def. 2 and transformation of spherical coordinates back to Cartesian coordinates, cf. (1.5)-(1.7), it can be shown [19] that each $Y_{l,m}$ corresponds to a homogeneous polynomial of degree l . Additionally, for fixed l , we have exactly $2l + 1$ spherical harmonics $Y_{l,m}$ of degree l , because $m = -l, \dots, l$. Since the set $\{Y_{l,m}\}$ is linearly independent and orthonormal, it is an orthonormal basis of $\text{Harm}_l(\Omega)$.

1.5 The Spherical Fourier Transform

Since spherical harmonics form a set of orthonormal functions on the unit sphere Ω , we may ask whether and in which sense this set is complete. A completeness property will then allow us to represent arbitrary functions as a (usually infinite) series of spherical harmonics. For numerical simulations we consider best approximations over a finite set of spherical harmonics, thus the question of approximation estimations arise.

Theorem 11. *The following statements are equivalent:*

- (i) $\{Y_{l,m}\}_{l=0,1,\dots, m=-l,\dots,l}$ is closed in $L^2(\Omega)$.
- (ii) The orthogonal expansion of any element $f \in L^2(\Omega)$ converges in norm to f , i.e.

$$\lim_{k \rightarrow \infty} \left\| f - \sum_{l=0}^k \sum_{m=-l}^l (f, Y_{l,m})_{L^2(\Omega)} Y_{l,m} \right\|_{L^2(\Omega)} = 0. \quad (1.78)$$

- (iii) For any $f \in L^2(\Omega)$, Parseval's identity

$$\|f\|_{L^2(\Omega)}^2 = (f, f)_{L^2(\Omega)} = \sum_{l=0}^{\infty} \sum_{m=-l}^l |(f, Y_{l,m})_{L^2(\Omega)}|^2 \quad (1.79)$$

holds.

- (iv) There is no strictly larger orthonormal system containing the orthonormal system $\{Y_{l,m}\}_{l=0,1,\dots, m=-l,\dots,l}$.

(v) The system $\{Y_{l,m}\}_{l=0,1,\dots, m=-l,\dots,l}$ has the completeness property. That is, $f \in L^2(\Omega)$ and $(f, Y_{l,m})_{L^2(\Omega)} = 0$ for all $l = 0, 1, \dots, m = -l, \dots, l$ implies $f = 0$.

(vi) An element $f \in L^2(\Omega)$ is determined uniquely by its orthogonal coefficients.

Proof. We refer to the literature (e.g. [11]) for a proof. \square

If we replace spherical harmonics $Y_{l,m}$ in Thm. 11 by circular harmonics $\cos(m\varphi)$ and $\sin(m\varphi)$ and readjust the sums appropriately, the well known results for Fourier series are obtained. Thus, Fourier series can be seen as a projection onto circular harmonics, so that it is natural to define the *spherical Fourier transform* as follows:

Definition 5 (Spherical Fourier Transform). The spherical Fourier transform $\mathcal{F}_\Omega : f \mapsto \mathcal{F}_\Omega\{f\}$ for $f \in L^1(\Omega)$ is defined by

$$(\mathcal{F}_\Omega\{f\})(l, m) = (f, Y_{l,m})_{L^2(\Omega)} = \int_0^\pi \int_0^{2\pi} f(\theta, \varphi) Y_{l,m}(\theta, \varphi) \sin \theta d\varphi d\theta . \quad (1.80)$$

For numerical applications, one is limited to approximations over a space of finite dimension, where asymptotic results as in Thm. 11 are insufficient for the choice of such a space. Additionally, one is interested in good error estimations so that in adaptive schemes the approximation space can be extended appropriately. Since the exact solution is usually unknown, information about the local smoothness of the true solution should be extractable from the approximate solution.

Theorem 12. Let $f \in C^{2k}(\Omega)$ with nonnegative integer k and its expansion into spherical harmonics be given as

$$f \sim \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{l,m} Y_{l,m} =: \sum_{l=0}^{\infty} Q_l , \quad (1.81)$$

where Q_l is a spherical harmonic of degree l and $\sum_{l=0}^{\infty} Q_l$ is called the condensed spherical harmonics expansion. Then for all $l > 0$ and $u \in \Omega$ there holds

$$|Q_l(u)| \leq \eta \|(\Delta^*)^k f\|_{L^2(\Omega)} l^{1/2-2k} , \quad (1.82)$$

where $(\Delta^*)^k$ denotes the k times iterated Beltrami operator and η is a constant.

Proof. A proof of this result can be found for example in [19]. \square

In particular, if we put $k = 1$ in (1.82), uniform convergence of the series (1.81) to f follows. Additionally, the rate of decay of $|Q_l|$ allows us to deduce the smoothness of the (for our purposes unknown) function f . The drawback of Thm. 12 is that only nonnegative integer values of k can be used. This can be resolved either by the use of interpolation theory, or by the use of a more abstract setting. In the following, we consider the latter and follow the exposition of Freedman et. al. [13].

We consider the linear space

$$\mathcal{A} = \{ \{A_{l,m}\} : A_{l,m} \in \mathbb{R}, l = 0, 1, \dots, m = -l, \dots, l \} \quad (1.83)$$

consisting of all sequences $\{A_{l,m}\}$ of real numbers $A_{l,m}$. Now, let $\{A_{l,m}\} \in \mathcal{A}$ be a sequence with $A_{l,m} \neq 0$ for all l and m .

Definition 6. The set $\mathcal{E} = \mathcal{E}(\{A_{l,m}\}; \Omega)$ is the set of all infinitely differentiable functions $f \in C^\infty(\Omega)$ such that

$$\sum_{n=0}^{\infty} \sum_{m=-l}^l |A_{l,m}|^2 (f, Y_{l,m})_{L^2(\Omega)}^2 < \infty . \quad (1.84)$$

We can define an inner product $(\cdot, \cdot)_{\mathcal{H}(\{A_{l,m}\}; \Omega)}$ on \mathcal{E} by

$$(f, g)_{\mathcal{H}(\{A_{l,m}\}; \Omega)} = \sum_{l=0}^{\infty} \sum_{m=-l}^l |A_{l,m}|^2 (f, Y_{l,m})_{L^2(\Omega)} (g, Y_{l,m})_{L^2(\Omega)} , \quad f, g \in \mathcal{E} . \quad (1.85)$$

The associated norm is then given by

$$\|f\|_{\mathcal{H}(\{A_{l,m}\}; \Omega)} = \sqrt{(f, f)_{\mathcal{H}(\{A_{l,m}\}; \Omega)}} . \quad (1.86)$$

A Hilbert space can now be obtained by completion:

Definition 7. The Sobolev space $\mathcal{H}(\{A_{l,m}\}; \Omega)$ is the completion of $\mathcal{E}(\{A_{l,m}\}; \Omega)$ under the norm $\|\cdot\|_{\mathcal{H}(\{A_{l,m}\}; \Omega)}$. $\mathcal{H}(\{A_{l,m}\}; \Omega)$ equipped with the inner product (1.85) is a Hilbert space.

The next step is to use the fact that the spherical harmonics $Y_{n,m}$ are eigenfunctions of the Beltrami operator, cf. Thm. 9:

$$\left(-\Delta^* + \frac{1}{4}\right)Y_{l,m} = \left(l(l+1) + \frac{1}{4}\right)Y_{l,m} = \left(l + \frac{1}{2}\right)^2 Y_{l,m} . \quad (1.87)$$

Thus, we formally have

$$\mathcal{F}_\Omega \left\{ \left(-\Delta^* + \frac{1}{4}\right)^{s/2} f \right\}(l, m) = \left(l + \frac{1}{2}\right)^s \mathcal{F}_\Omega \{f\}(l, m) \quad (1.88)$$

for $s \in \mathbb{R}$. This allows us to formally handle derivatives of functions by means of convergence properties of their spherical Fourier series. Thus, $(-\Delta^* + 1/4)^{s/2}$ is also referred to as *pseudodifferential operator* of order s .

Definition 8. We write

$$\mathcal{H}_s(\Omega) := \mathcal{H}\left(\left\{\left(l + \frac{1}{2}\right)^s\right\}; \Omega\right) \quad (1.89)$$

for the spherical Sobolev space of order s .

Using Parseval's identity, we (formally) have

$$\|f\|_{\mathcal{H}_s(\Omega)} = \|(-\Delta^* + 1/4)^{s/2} f\|_{L^2(\Omega)} , \quad (1.90)$$

and in particular, $\mathcal{H}_0(\Omega) = L^2(\Omega)$.

Similar as for Sobolev spaces defined over the whole space, the order parameter s is tightly connected to the smoothness of functions in $\mathcal{H}_s(\Omega)$:

Lemma 5. If $f \in \mathcal{H}_s(\Omega)$, where $s > k+1$, then f corresponds to a function of class $C^k(\Omega)$.

Proof. A proof can be found in the textbook of Freedman et. al. [13]. \square

With the formal apparatus at hand, we are now ready to estimate the error induced by truncated spherical Fourier series.

Theorem 13. For $f \in \mathcal{H}_s(\Omega)$, $s > 1$ there holds

$$\sup_{(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]} \left| f(\theta, \varphi) - \sum_{l=0}^N \sum_{m=-l}^l f_{l,m} Y_{l,m}(\theta, \varphi) \right| \leq \frac{C}{N^{s-1}} \|f\|_{\mathcal{H}_s(\Omega)} , \quad (1.91)$$

where

$$f_{l,m} := \mathcal{F}_\Omega\{f\}(l, m) . \quad (1.92)$$

Proof. Using the Cauchy-Schwarz inequality, we obtain (function arguments omitted)

$$\begin{aligned} \left| f - \sum_{l=0}^N \sum_{m=-l}^l f_{l,m} Y_{l,m} \right| &= \left| \sum_{l=N+1}^{\infty} \sum_{m=-l}^l f_{l,m} Y_{l,m} \right| \\ &\leq \left(\sum_{l=N+1}^{\infty} \sum_{m=-l}^l (f_{l,m})^2 (l + \frac{1}{2})^{2s} \right)^{1/2} \\ &\quad \times \left(\sum_{l=N+1}^{\infty} \sum_{m=-l}^l (Y_{l,m})^2 (l + \frac{1}{2})^{-2s} \right)^{1/2} \\ &= \left(\sum_{l=N+1}^{\infty} \sum_{m=-l}^l (Y_{l,m})^2 (l + \frac{1}{2})^{-2s} \right)^{1/2} \|f\|_{\mathcal{H}_s(\Omega)} . \end{aligned}$$

The remaining double sum can be reduced to a single sum by the use of the addition theorem (cf. (1.68)):

$$\sum_{l=N+1}^{\infty} \sum_{m=-l}^l (Y_{l,m})^2 (l + \frac{1}{2})^{-2s} = \frac{1}{2\pi} \sum_{l=N+1}^{\infty} (l + \frac{1}{2})^{-2s+1} . \quad (1.93)$$

The p -series can be estimated as follows:

$$\begin{aligned} \sum_{l=N+1}^{\infty} (l + \frac{1}{2})^{-2s+1} &\leq \sum_{l=N+1}^{\infty} \frac{1}{l^{2s-1}} \\ &\leq \frac{1}{N^{2s-1}} \sum_{l=N+1}^{\infty} \frac{1}{(l/N)^{2s-1}} \\ &\leq \frac{1}{N^{2s-1}} \sum_{l=1}^{\infty} \frac{N}{l^{2s-1}} \\ &= \frac{1}{N^{2s-2}} \sum_{l=1}^{\infty} \frac{1}{l^{2s-1}} = \frac{\zeta(2s-1)}{N^{2s-2}} \end{aligned}$$

where $\zeta(\cdot)$ denotes the Riemann zeta function. Thus, altogether we have just shown

$$\sup_{(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]} \left| f(\theta, \varphi) - \sum_{l=0}^N \sum_{m=-l}^l f_{l,m} Y_{l,m}(\theta, \varphi) \right| \leq \sqrt{\frac{\zeta(2s-1)}{2\pi}} \frac{1}{N^{s-1}} \|f\|_{\mathcal{H}_s(\Omega)} ,$$

hence the theorem holds with $C = \sqrt{\zeta(2s-1)}/2\pi$. \square

An estimate for $|Q_l|$ as in (1.82) can now be derived

Corollary 1. *Let $f \in \mathcal{H}_s(\Omega)$, $s > 1$, then for all $N > 1$ and $u \in \Omega$ there holds*

$$|Q_N(u)| \leq \eta N^{1-s} \|f\|_{\mathcal{H}_s(\Omega)} , \quad (1.94)$$

where η is a constant.

Proof. We find

$$\begin{aligned} |Q_N(u)| &= \left| \sum_{l=0}^N Q_l - \sum_{l=0}^{N-1} Q_l \right| = \left| \left(\sum_{l=0}^N Q_l - f \right) + \left(f - \sum_{l=0}^{N-1} Q_l \right) \right| \\ &\leq \left| f - \sum_{l=0}^N Q_l \right| + \left| f - \sum_{l=0}^{N-1} Q_l \right| \\ &\leq \left(\frac{C}{N^{s-1}} + \frac{C}{(N-1)^{s-1}} \right) \|f\|_{\mathcal{H}_s(\Omega)} \\ &= \frac{C + C(N/(N-1))^{s-1}}{N^{s-1}} \|f\|_{\mathcal{H}_s(\Omega)} \\ &\leq C(1 + 2^{s-1}) N^{1-s} \|f\|_{\mathcal{H}_s(\Omega)} . \end{aligned}$$

\square

Especially we find by setting $s = k + 1$ for continuously differentiable functions:

Corollary 2. *For $f \in C^k(\Omega)$ and $N > 1$, there holds*

$$|Q_N(u)| \leq \eta N^{-k} \|f\|_{\mathcal{H}_{k+1}(\Omega)} . \quad (1.95)$$

Chapter 2

The Boltzmann Transport Equation

In this chapter we deal with the mathematical and physical properties of the Boltzmann transport equation (BTE), which treats electrons as particles with a particular position $\mathbf{x}(t)$ and momentum $\mathbf{p}(t)$ at time t . In this equation, particle (electrons) are described by a distribution function $f(\mathbf{x}, \mathbf{p}, t)$, i.e. a probability density in the (\mathbf{x}, \mathbf{p}) -space. This is a classical kinetic model, since it defines both position and momentum of the particle. However, this contradicts Heisenberg's uncertainty principle

$$\Delta \mathbf{x} \Delta \mathbf{p} \geq \frac{\hbar}{2}, \quad (2.1)$$

so in principle it is impossible to determine both position and momentum with arbitrary precision. Despite its limitations for the accurate description of quantum mechanical effects, the semiclassical Boltzmann transport equation is seen as the best mathematical description of carrier transport and carrier distributions in semiconductor device modelling provided that the aforementioned quantum mechanical effects are negligible.

2.1 Derivation of the Transport Equation

Consider a single electron interpreted as a particle at position $\mathbf{x}(t) \in \mathbb{R}^3$ and momentum $\mathbf{p} \in \mathbb{R}^3$. These functions are a solution of Newton's laws of motion

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{p}}{dt} = \mathbf{F}, \quad t > 0 \quad (2.2)$$

supplemented with initial conditions $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{p}(0) = \mathbf{p}_0$. The force \mathbf{F} is given by $\mathbf{F} = q\mathbf{E}$, where the charge q is negative for electrons and positive for holes. \mathbf{E} denotes the effective electric field acting on the particle and is typically a superposition of externally applied fields and internal fields due to interactions among charges.

Instead of tracking a large number of electrons by Newton's laws, we describe the state of an electron by means of the distribution function $f(\mathbf{x}, \mathbf{p}, t)$. Then,

$$\int_B f(\mathbf{x}, \mathbf{p}, t) d\mathbf{x} d\mathbf{p} \quad (2.3)$$

gives the probability to find the particle at time t in the volume B in the position-momentum-space.

It is reasonable¹ to assume that $f(\mathbf{x}, \mathbf{p}, t)$ does not change along an trajectory given by $w(t; \mathbf{x}_0, \mathbf{p}_0)$, so that with the initial probability distribution $f_0(\mathbf{x}, \mathbf{p})$ we have

$$f(w(t; \mathbf{x}_0, \mathbf{p}_0), t) = f_0(\mathbf{x}, \mathbf{p}) \quad \forall t \geq 0. \quad (2.4)$$

Differentiation with respect to t and writing $\dot{\mathbf{x}}$ and $\dot{\mathbf{p}}$ for the time derivatives respectively yields

$$\frac{\partial f}{\partial t} + \dot{\mathbf{x}} \cdot \nabla_{\mathbf{x}} f + \dot{\mathbf{p}} \cdot \nabla_{\mathbf{p}} f = 0. \quad (2.5)$$

Using the relation $\mathbf{p} = m\mathbf{v} = m\dot{\mathbf{x}}$, we arrive with (2.2) at *Liouville's equation*

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{x}} f + \mathbf{F} \cdot \nabla_{\mathbf{p}} f = 0. \quad (2.6)$$

Liouville's equation considers only a single particle, whereas in semiconductors many particles interact simultaneously. Hence, we consider a generalisation to an ensemble of, say M , particles. The position and momentum vectors are then $3M$ -dimensional each. Additionally, the force field $\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_M)$ is a $3M$ -dimensional vector and in general each entry depends on the $6M$ position and momentum coordinates and on time t . The classical (ensemble) Liouville equation then reads

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{x}} f + \mathbf{F} \cdot \nabla_{\mathbf{p}} f = 0, \quad (2.7)$$

but is now posed for $\mathbf{x} \in \mathbb{R}^{3M}$, $\mathbf{p} \in \mathbb{R}^{3M}$. The joint particle density $f(\mathbf{x}, \mathbf{p}, t)$ can be interpreted similarly as in (2.3).

Up to now we described to motion of particles in vacuum. In semiconductors, however, there are interactions with the crystal lattice potential, which influences the motion of holes and electrons considerably. A classical description of such interactions cannot be used due to the small lattice constants, thus quantum mechanics has to be used to model the influence of the lattice potential. In the following we use the term 'electrons' instead of 'particles', even though the equations are valid for holes as well.

We will not go into all the details of quantum mechanics in semiconductors and refer to the literature [7, 29] for further information. One of the main results is that electrons occupy states within the energy bands. The energy of an electron in the l -th band is related with the wave vector $\mathbf{k} \in \mathbb{R}^3$ of the electron by $\varepsilon = \varepsilon_l(\mathbf{k})$ and the corresponding electron velocity is given by

$$\mathbf{v}_l(\mathbf{k}) = \frac{1}{\hbar} \nabla_{\mathbf{k}} \varepsilon_l(\mathbf{k}). \quad (2.8)$$

The energy bands are a material property and can only be obtained from experiments or sophisticated numerical simulations. Sufficiently accurate approximations can be found in the literature and are discussed for silicon in more detail in Sec. 2.3.

¹This can also be justified from the mathematics point of view, cf. Liouville's Theorem. [25]

After taking quantum mechanics into account, the equations of motion for the i -th electron are given by

$$\frac{d\mathbf{x}_i}{dt} = \mathbf{v}(k_i) , \quad \frac{d\mathbf{p}}{dt} = \hbar \frac{d\mathbf{k}}{dt} = \mathbf{F}_i , \quad t > 0 . \quad (2.9)$$

In the following it is more convenient to use the wave vector \mathbf{k} instead of the momentum \mathbf{p} as argument of the distribution function, thus we formally set $\tilde{f}(\mathbf{x}, \mathbf{k}, t) := f(\mathbf{x}, \hbar\mathbf{k}, t)$, but keep writing f instead of \tilde{f} . Writing $\mathbf{v}(\mathbf{k}) := (\mathbf{v}(\mathbf{k}_1), \dots, \mathbf{v}(\mathbf{k}_M))$, we obtain the *semiclassical electron ensemble Liouville equation* for the electron distribution function $f(\mathbf{x}, \mathbf{k}, t)$ as

$$\frac{\partial f}{\partial t} + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f + \frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f = 0 , \quad (2.10)$$

where again $\mathbf{x} \in \mathbb{R}^{3M}$, but $\mathbf{k}_i \in B$ for $i = 1, \dots, M$, where B denotes the *Brillouin zone* of the semiconductor. It is interesting to note that in the case of parabolic energy bands

$$\varepsilon(k) = \frac{\hbar^2 |\mathbf{k}|^2}{2m} , \quad (2.11)$$

we obtain $\mathbf{v} = \mathbf{p}/m = \hbar\mathbf{k}/m$ and the semiclassical Liouville equation reduces to its classical counterpart (2.7). Due to symmetry considerations at the construction of the Brillouin zone B it is natural to supplement (2.10) with periodic boundary conditions in \mathbf{k}_i , $i = 1, \dots, M$ on the boundary ∂B :

$$f(\mathbf{x}, \mathbf{k}_1, \dots, \mathbf{k}_i, \dots, \mathbf{k}_M, t) = f(\mathbf{x}, \mathbf{k}_1, \dots, -\mathbf{k}_i, \dots, \mathbf{k}_M, t) . \quad (2.12)$$

For numerical simulations, Liouville's equation cannot be solved for problems considered in practise due to its high dimensionality. In the following, we will consider the formal limit of the particle number M tending to infinity. Thus, we obtain a distribution function that describes an ensemble of many electrons while preserving the simpler mathematical structure (i.e. one space variable in \mathbb{R}^3 , one momentum variable in \mathbb{R}^3 and one time variable in \mathbb{R}) of the single-particle distribution function. Since there are several technical steps involved, only the cornerstones in the derivation will be given in the following. More details on the derivations can again be found in the literature (e.g. [30]).

We assume that the force acting on electron j can by superposition be written as

$$\mathbf{F}_j(\mathbf{x}, t) = q\mathbf{E}_{\text{ext}}(\mathbf{x}_j, t) + q \sum_{l=1, l \neq j}^M \mathbf{E}_{\text{int}}(\mathbf{x}_j, \mathbf{x}_l) , \quad (2.13)$$

where \mathbf{E}_{ext} denotes an externally applied field and \mathbf{E}_{int} is due to (long-range) two-particle interactions. The second main assumption is that the initial sub-ensemble particle density $f_I^{(d)}$ for d electrons can be factorised:

$$f_I^{(d)} = \prod_{j=1}^d P_I(\mathbf{x}_j, \mathbf{k}_j) , \quad d = 1, \dots, M-1 . \quad (2.14)$$

Thus, along with some further technical assumptions and symmetry considerations, it turns out that the sub-ensemble particle density can be factorised at all times $t > 0$, thus the

electron distribution for a large number of electrons can be reduced to the distribution function of a single electron, which fulfils the *Vlasov equation*

$$\frac{\partial f}{\partial t} + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f + \frac{q}{\hbar} \mathbf{E}_{\text{eff}} \cdot \nabla_{\mathbf{k}} f = 0 , \quad (2.15)$$

where the effective electric field is given by

$$\begin{aligned} \mathbf{E}_{\text{eff}}(\mathbf{x}, t) &= \mathbf{E}_{\text{ext}}(\mathbf{x}, t) + \int_{\mathbb{R}^3 \times B} f(\mathbf{y}, \mathbf{k}, t) \mathbf{E}_{\text{int}}(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{k} \\ &= \mathbf{E}_{\text{ext}}(\mathbf{x}, t) + \int_{\mathbb{R}^3} n(\mathbf{y}, t) \mathbf{E}_{\text{int}}(\mathbf{x}, \mathbf{y}) d\mathbf{y} \end{aligned} \quad (2.16)$$

with electron density $n(\cdot, \cdot)$, cf. Sec. 2.4. The macroscopic electric field (2.16) can readily be seen as the continuous extension of (2.13) taking the probability of finding electrons at a particular location into account. As for Liouville's equation, periodic boundary conditions as in (2.12) at the Brillouin zone boundaries are assumed.

For the common case that the interior electric field \mathbf{E}_{int} results from Coulomb forces in the form

$$\mathbf{E}_{\text{int}}(\mathbf{x}, \mathbf{y}) = -\frac{q}{4\pi\epsilon} \frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|^3} , \quad (2.17)$$

with permittivity ϵ , it can be shown [30] that the *Vlasov-Poisson system* is obtained:

$$\frac{\partial f}{\partial t} + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f - \frac{q}{\hbar} \nabla_{\mathbf{x}} \psi_{\text{eff}} \cdot \nabla_{\mathbf{k}} f = 0 , \quad (2.18)$$

$$\epsilon \Delta_{\mathbf{x}} \psi_{\text{eff}} = q(n - C) , \quad \mathbf{x} \in \mathbb{R}^3, \mathbf{k} \in B, t > 0 , \quad (2.19)$$

where ψ_{eff} is the (effective) electrostatic potential and C refers to the doping concentration.

So far we have only considered particle interactions through their induced (long-range) forces. However, if we consider electrons as (classical) particles, collisions among these particles occur, especially if the particle density is very high. At each collision event, the involved particles change their momentum, but not their location in space for a given point in time. We present a phenomenological derivation here, for more stringent derivations the reader is referred to the literature, e.g. [30].

In the derivation of Liouville's equation the only driving force for the electron was given by the externally applied electric field. The resulting trajectories are due to convection and we have assumed that there is no change of the probability density along trajectories, hence

$$\left(\frac{df}{dt} \right)_{\text{conv}} = 0 . \quad (2.20)$$

For reasons of symmetry of collision events it is postulated that the rate of change of f due to convection equals the rate of change due to collisions

$$\left(\frac{df}{dt} \right)_{\text{coll}} = \left(\frac{df}{dt} \right)_{\text{conv}} . \quad (2.21)$$

By convention, scattering is described by an operator $Q\{f\}$, so that we have finally arrived at the semiclassical *Boltzmann transport equation* (BTE)

$$\frac{\partial f}{\partial t} + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f + \frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f = Q\{f\} , \quad (2.22)$$

where the force is given as $\mathbf{F} = q\mathbf{E}_{\text{eff}}$. The collision operator $Q\{f\}$ is discussed in the next section and is in the most general form a nonlinear integral operator. Additionally, using Coulomb forces for long-range particle interactions, one obtains in analogy to the Vlasov-Poisson system the *Boltzmann-Poisson system*:

$$\frac{\partial f}{\partial t} + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f - \frac{q}{\hbar} \nabla_{\mathbf{x}} \psi_{\text{eff}} \cdot \nabla_{\mathbf{k}} f = Q\{f\} , \quad (2.23)$$

$$\epsilon \Delta_{\mathbf{x}} \psi_{\text{eff}} = q(n - C) , \quad \mathbf{x} \in \mathbb{R}^3, \mathbf{k} \in B, t > 0 . \quad (2.24)$$

A solution of the Boltzmann-Poisson system is often referred to as *self consistent solution* of the BTE. In addition to the nonlinearity of the scattering operator, this system is coupled in the position variable due to the self-consistent field $\mathbf{E}_{\text{eff}} = -\nabla_{\mathbf{x}} \psi_{\text{eff}}$. The initial conditions are given as for the Vlasov equation and the single-electron Liouville equation by

$$f(\mathbf{x}, \mathbf{k}, t) = f(\mathbf{x}, -\mathbf{k}, t) , \quad \mathbf{x} \in \mathbb{R}^3, \mathbf{k} \in \partial B, t > 0 , \quad (2.25)$$

$$f(\mathbf{x}, \mathbf{k}, 0) = f_0(\mathbf{x}, \mathbf{k}) , \quad \mathbf{x} \in \mathbb{R}^3, \mathbf{k} \in B . \quad (2.26)$$

2.2 Collision Events

Even though it is not apparent at first sight, several physical effects strongly depend on how the collision operator $Q\{f\}$ in the BTE (2.22) is modelled. In this section we consider different models for $Q\{f\}$, starting from the most general one. Clearly, only the most important scattering mechanisms can be discussed, so that we advise to consult the literature for more detailed information [29].

The first assumption on scattering events is that they occur instantaneously and only change the crystal momentum of the particles involved (cf. Fig. 2.1). The rate $S(\mathbf{x}, \mathbf{k} \rightarrow \mathbf{k}', t)$ at which a particle at (\mathbf{x}, t) changes its wave vector (and hence its momentum) from \mathbf{k} to \mathbf{k}' is proportional to two quantities:

- (i) The rate is proportional to the occupation probability $f(\mathbf{x}, \mathbf{k}, t)$. In terms of probabilities this means that a scattering event is more likely the higher the occupation probability for the initial state given by \mathbf{k} is.
- (ii) Electrons consider the *Pauli principle*: The final state \mathbf{k}' has to be vacant, thus the scattering rate is assumed to be proportional to $1 - f(\mathbf{x}, \mathbf{k}', t)$.

Altogether the rate of transition events from state \mathbf{k} to state \mathbf{k}' thus is

$$S(\mathbf{x}, \mathbf{k} \rightarrow \mathbf{k}', t) = s(\mathbf{x}, \mathbf{k}, \mathbf{k}') f(\mathbf{x}, \mathbf{k}, t) (1 - f(\mathbf{x}, \mathbf{k}', t)) , \quad (2.27)$$

where $s(\mathbf{x}, \mathbf{k}, \mathbf{k}')$ denotes the *scattering rate*. The rate of change of the distribution function f for a particular wave vector \mathbf{k} is then the sum over all differences of in-scattering and out-scattering rates, taken over all possible wave vectors (Bloch states):

$$\begin{aligned} (Q\{f\})(\mathbf{x}, \mathbf{k}, t) &= \int_B S(\mathbf{x}, \mathbf{k}' \rightarrow \mathbf{k}, t) - S(\mathbf{x}, \mathbf{k} \rightarrow \mathbf{k}', t) d\mathbf{k}' \\ &= \int_B s(\mathbf{x}, \mathbf{k}', \mathbf{k}) f(\mathbf{x}, \mathbf{k}', t) (1 - f(\mathbf{x}, \mathbf{k}, t)) d\mathbf{k}' \\ &\quad - \int_B s(\mathbf{x}, \mathbf{k}, \mathbf{k}') f(\mathbf{x}, \mathbf{k}, t) (1 - f(\mathbf{x}, \mathbf{k}', t)) d\mathbf{k}' . \end{aligned} \quad (2.28)$$

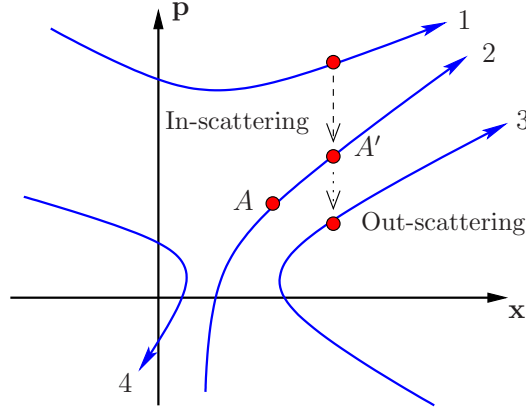


Figure 2.1: Illustration of trajectories in position-momentum space. Carriers move along a trajectory according to Newton's Laws. Occasionally they scatter to another trajectory. Scattering instantly changes the carrier's momentum, but does not affect its position.

The first integral is often referred to as *in-scattering term*, while the second is called *out-scattering term*, cf. Fig. 2.1. This general form of the collision operator induces a nonlocal quadratic nonlinearity into the BTE (2.22).

At first sight one might be tempted to assume symmetry of the scattering rates with respect to the wave vectors: $s(\mathbf{x}, \mathbf{k}', \mathbf{k}) \stackrel{?}{=} s(\mathbf{x}, \mathbf{k}, \mathbf{k}')$. However, this is not true in general, but a different statement holds in thermal equilibrium: By the *principle of detailed balance*, the local scattering probabilities vanish, hence

$$s(\mathbf{x}, \mathbf{k}', \mathbf{k}) f_{\text{eq}}(\mathbf{x}, \mathbf{k}', t) (1 - f_{\text{eq}}(\mathbf{x}, \mathbf{k}, t)) - s(\mathbf{x}, \mathbf{k}, \mathbf{k}') f_{\text{eq}}(\mathbf{x}, \mathbf{k}, t) (1 - f_{\text{eq}}(\mathbf{x}, \mathbf{k}', t)) = 0. \quad (2.29)$$

Since the equilibrium occupation number density is given by the Fermi-Dirac distribution

$$f_{\text{eq}}(\mathbf{x}, \mathbf{k}) = \frac{1}{1 + \exp\left(\frac{\varepsilon(\mathbf{k}) - \varepsilon_F}{k_B T}\right)}, \quad (2.30)$$

we have

$$1 - f_{\text{eq}}(\mathbf{x}, \mathbf{k}) = \frac{\exp\left(\frac{\varepsilon(\mathbf{k}) - \varepsilon_F}{k_B T}\right)}{1 + \exp\left(\frac{\varepsilon(\mathbf{k}) - \varepsilon_F}{k_B T}\right)}. \quad (2.31)$$

Substitution of (2.30) and (2.31) into (2.29) and cancelling the common denominator leads to

$$\begin{aligned} s(\mathbf{x}, \mathbf{k}', \mathbf{k}) \exp\left(\frac{\varepsilon(\mathbf{k}) - \varepsilon_F}{k_B T}\right) &= s(\mathbf{x}, \mathbf{k}, \mathbf{k}') \exp\left(\frac{\varepsilon(\mathbf{k}') - \varepsilon_F}{k_B T}\right) \\ \iff \frac{s(\mathbf{x}, \mathbf{k}, \mathbf{k}')}{s(\mathbf{x}, \mathbf{k}', \mathbf{k})} &= \exp\left(\frac{\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}')}{k_B T}\right). \end{aligned} \quad (2.32)$$

The scattering rate $s(\cdot, \cdot, \cdot)$ is determined from the type of the scattering mechanism. We briefly discuss the most important ones:

- *Phonon scattering:* The atoms in the crystal lattice vibrate at nonzero temperature about their fixed equilibrium locations. These vibrations are quantised and the quantum of lattice vibrations is called *phonon*, carrying the energy $\hbar\omega_{\text{phon}}$. Thus an electron that collides with a phonon either increases or decreases its energy by $\hbar\omega_{\text{phon}}$. With the phonon occupation number N_{phon} given by the Bose-Einstein statistics

$$N_{\text{phon}} = \frac{1}{\exp\left(\frac{\hbar\omega_{\text{phon}}}{k_B T}\right) - 1} , \quad (2.33)$$

the scattering rate for initial state \mathbf{k} and final state \mathbf{k}' can be written as

$$s_{\text{phon}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') = \sigma_{\text{phon}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') \left[N_{\text{phon}} \delta(\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}') + \hbar\omega_{\text{phon}}) + (1 + N_{\text{phon}}) \delta(\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}') - \hbar\omega_{\text{phon}}) \right] , \quad (2.34)$$

where $\sigma_{\text{phon}}(\mathbf{x}, \mathbf{k}, \mathbf{k}')$ is symmetric in \mathbf{k} and \mathbf{k}' and $\delta(\cdot)$ denotes the delta distribution. From the physics point of view one can additionally distinguish between *acoustic phonons* and *optical phonons*, but in both cases the scattering rate is given by (2.34). From the mathematics point of view it has to be emphasised that (2.34) is highly non-smooth, which further complicates numerical solutions of the BTE.

- *Ionised impurity scattering:* Dopants in a semiconductor are incorporated into the lattice as fixed charges. Since electrons also carry a charge, their trajectories are influenced by these fixed charges due to dopants, changing their momentum. This scattering process is elastic, i.e. the electron energy $\varepsilon(\mathbf{k}')$ after the collision equals the electron energy $\varepsilon(\mathbf{k})$ before interaction. Thus, the scattering rate s_{imp} is given similarly to (2.34) by

$$s_{\text{imp}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') = \sigma_{\text{imp}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') \delta(\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}')) , \quad (2.35)$$

where $\sigma_{\text{imp}}(\mathbf{x}, \mathbf{k}, \mathbf{k}')$ is symmetric in \mathbf{k} and \mathbf{k}' . Due to the symmetry of σ_{imp} there holds

$$\begin{aligned} s_{\text{imp}}(\mathbf{x}, \mathbf{k}', \mathbf{k}) f(\mathbf{x}, \mathbf{k}', t) f(\mathbf{x}, \mathbf{k}, t) - s_{\text{imp}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') f(\mathbf{x}, \mathbf{k}, t) f(\mathbf{x}, \mathbf{k}', t) = \\ = \sigma_{\text{imp}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') \delta(\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}')) [f(\mathbf{x}, \mathbf{k}', t) f(\mathbf{x}, \mathbf{k}, t) - f(\mathbf{x}, \mathbf{k}, t) f(\mathbf{x}, \mathbf{k}', t)] = 0 , \end{aligned}$$

hence the collision operator (2.28) can be simplified to

$$(Q_{\text{imp}}\{f\})(\mathbf{x}, \mathbf{k}, t) = \int_B \sigma_{\text{imp}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') \delta(\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}')) [f(\mathbf{x}, \mathbf{k}', t) - f(\mathbf{x}, \mathbf{k}, t)] d\mathbf{k}' . \quad (2.36)$$

- *Carrier-carrier scattering:* Apart from electron-phonon and electron-impurity interaction, electron-electron interaction is one of the most important scattering mechanisms. Since both electrons and holes are carriers, also electron-hole scattering can be considered, but only electron-electron scattering is considered here. For two electrons in Bloch states \mathbf{k} and $\tilde{\mathbf{k}}$, energy has to be conserved, hence

$$\varepsilon(\mathbf{k}) + \varepsilon(\tilde{\mathbf{k}}) = \varepsilon(\mathbf{k}') + \varepsilon(\tilde{\mathbf{k}}') , \quad (2.37)$$

where \mathbf{k}' and $\tilde{\mathbf{k}}'$ denote the Bloch states after the scattering event. The scattering rate can now be written as

$$s_{ee}(\mathbf{x}, \mathbf{k}, \mathbf{k}', \tilde{\mathbf{k}}, \tilde{\mathbf{k}}') = \sigma_{ee}(\mathbf{x}, \mathbf{k}, \mathbf{k}', \tilde{\mathbf{k}}, \tilde{\mathbf{k}}') \delta(\varepsilon(\mathbf{k}) + \varepsilon(\tilde{\mathbf{k}}) - \varepsilon(\mathbf{k}') - \varepsilon(\tilde{\mathbf{k}}')) , \quad (2.38)$$

with parameter $\sigma_{ee}(\mathbf{x}, \mathbf{k}, \mathbf{k}', \tilde{\mathbf{k}}, \tilde{\mathbf{k}}')$. We obtain the collision operator by taking into account that both initial states have to be occupied and that both final states have to be vacant:

$$\begin{aligned} (Q_{ee}\{f\})(\mathbf{x}, \mathbf{k}, t) = & \int_B s_{ee}(\mathbf{x}, \mathbf{k}, \mathbf{k}', \tilde{\mathbf{k}}, \tilde{\mathbf{k}}') \\ & \left[f(\mathbf{x}, \mathbf{k}', t) f(\mathbf{x}, \tilde{\mathbf{k}}', t) (1 - f(\mathbf{x}, \mathbf{k}, t)) (1 - f(\mathbf{x}, \tilde{\mathbf{k}}, t)) - \right. \\ & \left. f(\mathbf{x}, \mathbf{k}, t) f(\mathbf{x}, \tilde{\mathbf{k}}, t) (1 - f(\mathbf{x}, \mathbf{k}', t)) (1 - f(\mathbf{x}, \tilde{\mathbf{k}}', t)) \right] d\mathbf{k}' d\tilde{\mathbf{k}} d\tilde{\mathbf{k}}' . \end{aligned} \quad (2.39)$$

We observe that this collision operator has a nonlocal nonlinearity of fourth order.

Typically the scattering mechanisms are assumed to be independent from each other. The full collision operator is then taken as the sum of all individual scattering mechanisms:

$$Q\{f\} = Q_{\text{phon}}\{f\} + Q_{\text{imp}}\{f\} + Q_{ee}\{f\} . \quad (2.40)$$

Such a combination of individual scattering effects is often computationally too complicated or too expensive, thus simplifications are desired. We will discuss two of the most commonly used approximations:

- *Low density approximation:* Assuming that $0 \leq f(\mathbf{x}, \mathbf{k}, t) \ll 1$ and hence $1 - f(\mathbf{x}, \mathbf{k}, t) \approx 1$ almost everywhere, the collision operator as given in (2.28) simplifies using (2.32) to

$$Q_0(f)(\mathbf{x}, \mathbf{k}, t) \simeq \int_B \sigma(\mathbf{x}, \mathbf{k}, \mathbf{k}') \left[\exp\left(\frac{\varepsilon(\mathbf{k}')}{k_B T}\right) f(\mathbf{x}, \mathbf{k}', t) - \exp\left(\frac{\varepsilon(\mathbf{k})}{k_B T}\right) f(\mathbf{x}, \mathbf{k}, t) \right] d\mathbf{k}' , \quad (2.41)$$

where $\sigma(\mathbf{x}, \mathbf{k}, \mathbf{k}') = s(\mathbf{x}, \mathbf{k}', \mathbf{k}) \exp(-\varepsilon(\mathbf{k}')/(k_B T))$ is called *collision cross-section*. It is important to note that the collision cross-section is symmetric in \mathbf{k} and \mathbf{k}' because of (2.32). With this approximation the collision operator becomes linear, but the unknown distribution function $f(\mathbf{x}, \mathbf{k}, t)$ still appears in the integrand, so that the Boltzmann transport equation is still an integro-differential equation.

- *Relaxation time approximation:* In the case that the initial distribution is close to a Maxwell distribution, one can approximate

$$f(\mathbf{x}, \mathbf{k}', t) \approx n(\mathbf{x}) \exp\left(\frac{\varepsilon(\mathbf{k}')}{k_B T}\right) \quad (2.42)$$

in (2.41) and obtain

$$Q_\tau\{f\}(\mathbf{x}, \mathbf{k}, t) = -\frac{1}{\tau(\mathbf{x}, \mathbf{k})} \left[f(\mathbf{x}, \mathbf{k}, t) - \exp\left(\frac{\varepsilon(\mathbf{k})}{k_B T}\right) n(\mathbf{x}) \right] , \quad (2.43)$$

where

$$\tau(\mathbf{x}, \mathbf{k}) = \left(\int_B s(\mathbf{x}, \mathbf{k}, \mathbf{k}') d\mathbf{k}' \right)^{-1} \quad (2.44)$$

is called *relaxation time* and

$$n(x) = \int_B f_0(\mathbf{x}, \mathbf{k}') d\mathbf{k}' \quad (2.45)$$

is some given density. While the low density approximation removes the nonlinearity of the collision operator, the relaxation time approximation additionally transforms the integro-differential Boltzmann equation into a (hyperbolic) partial differential equation.

Although these simplifications of the collision operator ease numerical simulations considerably, an accurate simulation of several physical effects requires the full nonlinearity of the collision operator. In particular, the relaxation time approximation cannot be used if the equilibrium distribution is unknown. The low density approximation cannot be used in highly doped or degenerate semiconductors, where many electrons populate the conduction band and thus $1 - f(\mathbf{x}, \mathbf{k}, t) \not\approx 1$.

2.3 Multi-Band Structures

So far we have considered the BTE for a single energy band only. For predictive device simulation, the band structure of the semiconductor has to be closely approximated, thus more complicated energy band structures than the simplistic parabolic single-band structure have to be considered. In this section we first discuss the parabolic band approximation, then cover non-parabolicity and finally extended the scheme to cover multiple (spherical) bands. Ultimately, the most commonly used analytical multi-band structure is discussed and we hint at the necessary extensions to the BTE to cover multi-band effects.

In an idealised setting such as the infinitely deep quantum well, the solution of Schrödinger's equation yields a parabolic dependence of the energy on the wave vector:

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2 |\mathbf{k}|^2}{2m^*}, \quad (2.46)$$

where \hbar is the scaled Planck constant and m^* is the effective mass. Such a quadratic relationship is termed *parabolic band approximation*.

In a crystal lattice, the potential distribution deviates from such ideal quantum wells, hence the parabolic band approximation is indeed only an approximation of the real band structure. Nevertheless, the band structures observed in practise show only small deviations from the ideal parabolic shape. Empirical models suggest an analytic band model of the form

$$\gamma(\varepsilon) = \frac{\hbar^2 \mathbf{k}^2}{2m^*}, \quad (2.47)$$

where typically

$$\gamma(\varepsilon) = \varepsilon(1 + \alpha\varepsilon) \quad (2.48)$$

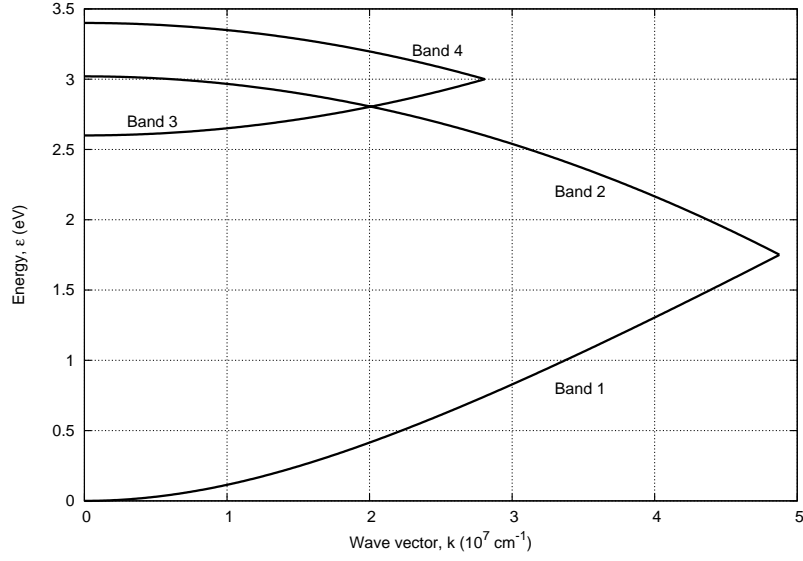


Figure 2.2: Multi-band structure of silicon. Bands 1 and 3 are electron like (increasing density of states), bands 2 and 4 are hole-like (decreasing density of states).

is used. The parameter α is called *non-parabolicity factor*; in case $\alpha = 0$ we readily fall back to (2.46). One can think of more complicated dependencies of the energy on the wave vector, but (2.48) already provides very good results in the low-energy regime.

For silicon, four energy bands have been proposed by Brunetti et. al. [8] given by

$$\varepsilon + \alpha\varepsilon^2 = \frac{\hbar^2 \mathbf{k}^2}{2m^{*(1)}} , \quad (\text{Band 1}) \quad (2.49)$$

$$\varepsilon = \varepsilon_{\max}^{(2)} - \frac{\hbar^2 \mathbf{k}^2}{2m^{*(2)}} , \quad (\text{Band 2}) \quad (2.50)$$

$$\varepsilon = \varepsilon_{\min}^{(3)} + \frac{\hbar^2 \mathbf{k}^2}{2m^{*(3)}} , \quad (\text{Band 3}) \quad (2.51)$$

$$\varepsilon = \varepsilon_{\max}^{(4)} - \frac{\hbar^2 \mathbf{k}^2}{2m^{*(4)}} , \quad (\text{Band 4}) \quad (2.52)$$

with energy limits $\varepsilon_{\max}^{(2)}$, $\varepsilon_{\min}^{(3)}$ and $\varepsilon_{\max}^{(4)}$ of the respective bands. The specific form of each band is

$$\gamma^{(1)}(\varepsilon) = \varepsilon + \alpha\varepsilon^2 , \quad (\text{Band 1}) \quad (2.53)$$

$$\gamma^{(2)}(\varepsilon) = \varepsilon_{\max}^{(2)} - \varepsilon , \quad (\text{Band 2}) \quad (2.54)$$

$$\gamma^{(3)}(\varepsilon) = \varepsilon - \varepsilon_{\min}^{(3)} , \quad (\text{Band 3}) \quad (2.55)$$

$$\gamma^{(4)}(\varepsilon) = \varepsilon_{\max}^{(4)} - \varepsilon , \quad (\text{Band 4}) \quad (2.56)$$

and we obtain the relationships between the energy and the wave vector in a compact form known as *dispersion relation* as

$$\gamma^{(\nu)}(\varepsilon) = \frac{\hbar^2 \mathbf{k}^2}{2m^{*(\nu)}} , \quad \nu = 1, 2, 3, 4 , \quad (2.57)$$

Band	m^*/m_0	ε_{\min}	ε_{\max}	Z	α
1	0.320	0	1.75	6	0.35
2	0.712	1.75	3.02	6	0
3	0.750	2.60	3.00	12	0
4	0.750	3.00	3.40	12	0

Table 2.1: Parameters for the analytical band structure proposed by Brunetti et. al. [8]. m_0 denotes the electron mass. Each band extends from ε_{\min} to ε_{\max} .

shown in Fig. 2.2.

For later use in the next section we briefly mention that the total density of states $g(\varepsilon)$, that is the number of states per unit volume in energy space for a particular energy, is computed from the individual density of states for each energy band by a weighted sum:

$$g(\varepsilon) = Z^{(1)}g^{(1)}(\varepsilon) + Z^{(2)}g^{(2)}(\varepsilon) + Z^{(3)}g^{(3)}(\varepsilon) + Z^{(4)}g^{(4)}(\varepsilon) , \quad (2.58)$$

where the *band multiplicities* $Z^{(\nu)}$, $\nu = 1, 2, 3, 4$ account for the number of equivalent symmetrical bands of the ν -th band. These band multiplicities and other band structure parameters can be found in Tab. 2.1.

For a multi-band structure, the BTE has in principle to be solved in each band for the band distribution function $f^{(\nu)}$ separately and the scattering operator has to consider scattering events between these bands. For the analytical band structure presented here, this would result in $6 + 6 + 12 + 12 = 36$ coupled equations. However, since electrons in equivalent valleys of the same band behave identically, one equation per band is sufficient, thus four equations remain. Additionally, it is possible to treat the first and the second band as a single band by

$$\gamma^{(1,2)}(\varepsilon) = \begin{cases} \varepsilon(1 + \alpha\varepsilon) , & 0 \text{ eV} \leq \varepsilon \leq 1.75 \text{ eV} , \\ \varepsilon_{\max}^{(2)} - \varepsilon , & 1.75 \text{ eV} < \varepsilon \leq 3.02 \text{ eV} . \end{cases} \quad (2.59)$$

Similarly, the third and forth band can be collected to obtain $\gamma^{(3,4)}$. If only small fields occur, it is typically sufficient to consider the first and second band only, so that only a single distribution function has to be computed [16, 48]. We note that $\gamma^{(1,2)}(\varepsilon)$ is neither differentiable nor even continuous at $\varepsilon = 1.75$ eV in a classical sense, thus in the following derivatives of $\gamma^{(1,2)}(\varepsilon)$ have to be understood in a weak sense.

For the remainder of this thesis we omit the band index ν in all expressions to avoid notational clutter. Hence, whenever band-specific quantities such as the distribution function occur, the band index ν is implicitly included but not written explicitly.

2.4 Average Quantities and Moments

A solution of the BTE yields the distribution function $f(\mathbf{x}, \mathbf{k}, t)$. For engineering purposes, macroscopic quantities such as the carrier density and the current density in the device are of interest. The computation of these macroscopic quantities from the distribution function is described in this section.

We start with the calculation of the *carrier concentration* $N(V, t)$ in a small volume V . On average, $N(V, t)$ is proportional (for brevity we assume the proportionality factor to be unity) to the sum over all occupation probabilities of the energy states in that volume:

$$N(V, t) = \int_V \sum_{\text{all states at } \mathbf{x}} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{x} . \quad (2.60)$$

An evaluation of this sum is avoided by taking the formal limit of a continuous distribution of states:

$$N(V, t) = \frac{2}{(2\pi)^3} \int_V \int_{\mathbb{R}^3} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} d\mathbf{x} , \quad (2.61)$$

where the prefactor $1/(2\pi)^3$ arises due to different metrics in the \mathbf{k} -space and the \mathbf{x} -space and the additional factor of two is due to the fact that each state can be occupied by two electrons with different spin. The *carrier density* $n(\mathbf{x}, t)$ is then obtained in the limit

$$n(\mathbf{x}, t) = \lim_{|V| \rightarrow 0} \frac{N(V, t)}{|V|} = \frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} . \quad (2.62)$$

In case that spherical coordinates $(\varepsilon, \theta, \varphi)$ are used in \mathbf{k} -space, integrals over the whole \mathbf{k} -space have to be transformed appropriately. For the carrier density as given in (2.62), a change of coordinates using $k = |\mathbf{k}|$ yields

$$\begin{aligned} \frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} &= \frac{2}{(2\pi)^3} \int_0^\infty \int_0^\pi \int_0^{2\pi} f(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t) k^2 \sin \theta \frac{\partial k}{\partial \varepsilon} d\varphi d\theta d\varepsilon \\ &= \frac{2}{(2\pi)^3} \int_0^\infty \int_\Omega f(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t) k^2 \frac{\partial k}{\partial \varepsilon} d\Omega d\varepsilon , \end{aligned} \quad (2.63)$$

where Ω denotes the unit sphere and $d\Omega = \sin \theta d\varphi d\theta$. We define the *generalised density of states* as

$$g(\varepsilon) := \frac{k^2}{(2\pi)^3} \frac{\partial k}{\partial \varepsilon} , \quad (2.64)$$

so that we can now write

$$\frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} = 2 \int_0^\infty \int_\Omega f(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t) g(\varepsilon, \theta, \varphi) d\Omega d\varepsilon . \quad (2.65)$$

If a spherical relation (2.57) between energy and wave vector holds, the triple integral can be further simplified to a single integral over energy: Let $f(\mathbf{x}, \mathbf{k}, t)$ depend on the modulus $k = |\mathbf{k}|$ only (or in this case equivalently depend on ε only, but not on the angles θ and φ), then we immediately obtain from (2.57)

$$k = k(\varepsilon) = \sqrt{\frac{2m^*\gamma(\varepsilon)}{\hbar^2}} \quad (2.66)$$

$$\implies \frac{dk}{d\varepsilon} = \sqrt{\frac{\hbar^2}{2m^*\gamma(\varepsilon)}} \gamma'(\varepsilon) \frac{m^*}{\hbar^2} , \quad (2.67)$$

where $\gamma'(\varepsilon)$ denotes the derivative of $\gamma(\varepsilon)$ with respect to ε . Consequently, $g(\varepsilon, \theta, \varphi) = g(\varepsilon)$ and we obtain

$$2 \int_0^\infty \int_\Omega f(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t) g(\varepsilon, \theta, \varphi) d\Omega d\varepsilon = 8\pi \int_0^\infty f(\mathbf{x}, \mathbf{k}(\varepsilon), t) g(\varepsilon) d\varepsilon \quad (2.68)$$

The *physical density of states* is then be found as

$$g_{\text{phys}}(\varepsilon) = 8\pi g(\varepsilon) = \frac{\sqrt{2}(m^*)^{3/2}}{\pi^2 \hbar^3} \sqrt{\gamma(\varepsilon)} \gamma'(\varepsilon) \quad (2.69)$$

and we thus have for spherical bands the relationship

$$\frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} F(k) d\mathbf{k} = \int_0^\infty F(k(\varepsilon)) g_{\text{phys}}(\varepsilon) d\varepsilon \quad (2.70)$$

for any function F that depends on the modulus k of the wave vector \mathbf{k} only. The carrier density (2.62) for spherical bands can then be written in compact form as

$$n(\mathbf{x}, t) = \int_0^\infty f(\mathbf{x}, \mathbf{k}(\varepsilon), t) g_{\text{phys}}(\varepsilon) d\varepsilon. \quad (2.71)$$

Macroscopic quantities depend on the location and time, thus the additional wave vector information in the distribution function has to be interpreted appropriately. The concept used for the calculation of the carrier density in (2.62) can be generalised for a small volume V and a weight function W as

$$\langle NW \rangle := \int_V \sum_{\text{all states at } \mathbf{x}} W(\mathbf{x}, \mathbf{k}, t) f(\mathbf{x}, \mathbf{k}, t) d\mathbf{x} \quad (2.72)$$

$$\rightarrow \frac{2}{(2\pi)^3} \int_V \int_{\mathbb{R}^3} W(\mathbf{x}, \mathbf{k}, t) f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} d\mathbf{x}, \quad (2.73)$$

where N denotes the number of electrons in V and we have taken the limit of a continuous distribution of states. Again, an average quantity in the limit $|V| \rightarrow 0$ can be obtained as

$$\langle nW \rangle := \lim_{|V| \rightarrow 0} \frac{\langle NW \rangle}{|V|}. \quad (2.74)$$

Since $\langle nW \rangle$ still depends on the carrier density, a normalised quantity for each particle is obtained as

$$\langle W \rangle = \frac{\langle nW \rangle}{\langle n \rangle}. \quad (2.75)$$

Typically, the weight function is chosen such that it depends on the wave vector only ($W = W(\mathbf{k})$). In case $W_r(\mathbf{k}) := \mathbf{k}^r$, where powers of \mathbf{k} are defined by means of the vector scalar product, the quantities $\langle nW_r \rangle$ and $\langle W_r \rangle$ are called r -th moment and r -th normalised moment of $f(\mathbf{x}, \mathbf{k}, t)$ respectively. Note that for even values of r the resulting moments are scalar valued, while they are vector valued for odd r ; the integrals then have to be taken component-wise.

With this new notation, the mean electron velocity is $\langle \mathbf{u}_g \rangle$, where

$$\mathbf{u}_g(\mathbf{k}) = \frac{1}{\hbar} \nabla_{\mathbf{k}} \varepsilon(\mathbf{k}) \quad (2.76)$$

is the group velocity. For spherical bands as in (2.57) it can be expressed by means of the chain rule:

$$\frac{d\gamma}{dk_i} = \frac{d\gamma}{d\varepsilon} \frac{d\varepsilon}{dk_i} , \quad i = 1, 2, 3 ,$$

with $\mathbf{k} = (k_1, k_2, k_3)$ to find

$$\begin{aligned} \mathbf{u}_g(\mathbf{k}) &= \frac{\hbar}{m^* \gamma'(\varepsilon)} \mathbf{k} \\ \Rightarrow \mathbf{u}_g(\varepsilon) &= \frac{\hbar}{m^* \gamma'(\varepsilon)} \sqrt{\frac{2m^* \gamma(\varepsilon)}{\hbar^2}} \mathbf{e}_\varepsilon = \frac{1}{\gamma'(\varepsilon)} \sqrt{\frac{2\gamma(\varepsilon)}{m^*}} \mathbf{e}_\varepsilon , \end{aligned}$$

where $\gamma'(\varepsilon)$ denotes the derivative with respect to ε and \mathbf{e}_ε the radial unit vector. With the average group velocity at hand, the current density \mathbf{j} can be directly obtained from

$$\mathbf{j} = \langle n\mathbf{u} \rangle . \quad (2.77)$$

Another important quantity of interest is the average energy, which is in our notation given by $\langle \varepsilon \rangle$. If we assume parabolic bands, then we immediately obtain from (2.46) that the average energy is proportional to the second moment. More precisely,

$$\langle \varepsilon \rangle = \frac{\hbar^2}{2m^*} \langle W_2 \rangle . \quad (2.78)$$

Chapter 3

Spherical Projection of the Boltzmann Transport Equation

In Chapter 1 it has been shown that spherical harmonics are orthonormal and complete on the sphere (Thm. 8 and Thm. 11). In Chapter 2, a distribution function for electrons was introduced, whose arguments are the spatial location, the wave vector and time. Thus, governing equation for the distribution function, the Boltzmann transport equation (BTE), is seven-dimensional. However, since the electron energy depends in good approximation on the modulus of the wave vector only, this spherical symmetry can be exploited to reduce the three-dimensional momentum space to a one-dimensional energy space. This idea has led to an expansion of the wave vector dependence in spherical harmonics for the purpose of numerical simulation in the early 1990s [15,17]. For analytical investigations, first spherical harmonics expansions (SHE) of the distribution function already date back to the 1950s.

3.1 The Spherical Harmonics Expansion Equations

In this section we aim to express the distribution function as a series of spherical harmonics in the form

$$f(\mathbf{x}, \mathbf{k}, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{l,m}(\mathbf{x}, \varepsilon, t) Y_{l,m}(\theta, \varphi) , \quad (3.1)$$

Only a finite number of terms can be considered in a numerical simulation, thus

$$f(\mathbf{x}, \mathbf{k}, t) \approx \sum_{l=0}^L \sum_{m=-l}^l f_{l,m}(\mathbf{x}, \varepsilon, t) Y_{l,m}(\theta, \varphi) , \quad (3.2)$$

which is for smooth f fully justified when looking at Thm. 13 and taking the finite precision of computers into account. In first publications only expansions up to first order have been considered. In spite of the low order expansion, satisfactory results from the physics point of view have been obtained in the low field regime [15,16,28,43,45–47,49].

Before any numerical simulation can be carried out, equations for the unknown coefficients $f_{l,m}(\mathbf{x}, \varepsilon, t)$ need to be derived from the BTE. There are two ways to proceed: One is to substitute (3.2) into the BTE and balance terms of similar order of magnitude. Such

an approach was used for example by Ventura, Gnudi and Baccarani [15, 16, 48, 49], which suffers from a lot of book-keeping and requires good reasoning on the balancing of terms. Additionally, this approach is limited to low-order expansions. We follow the second approach, where the BTE is projected onto each spherical harmonic and from this conditions for the expansion coefficients are deduced [20, 21, 24, 27, 33]

Before we proceed with a projection of the full BTE, we recall (2.65) for an integration over surfaces of constant energy ε :

$$\begin{aligned} \frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} \delta(\varepsilon - \varepsilon(\mathbf{k})) Y_{l,m}(\theta(\mathbf{k}), \varphi(\mathbf{k})) f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} = \\ = 2 \int_{\Omega} Y_{l,m}(\theta, \varphi) f(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t) g(\varepsilon, \theta, \varphi) d\Omega . \end{aligned} \quad (3.3)$$

Here, $\delta(\cdot)$ denotes the delta distribution. Note that since $Y_{0,0}$ is constant, the carrier density (2.71) can be obtained from an expansion of f into spherical harmonics in case of spherical bands as

$$n(\mathbf{x}, t) = \frac{2}{Y_{0,0}} \int_0^\infty f_{0,0}(\mathbf{x}, \varepsilon, t) g(\varepsilon) d\varepsilon . \quad (3.4)$$

Consequently, a projection of the distribution function onto $Y_{l,m}(\theta, \varphi)$ reads

$$f_{l,m}(\mathbf{x}, \varepsilon, t) = 2 \int_{\Omega} Y_{l,m}(\theta, \varphi) f(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t) g(\varepsilon, \theta, \varphi) d\Omega \quad (3.5)$$

$$=: \int_{\Omega} Y_{l,m}(\theta, \varphi) \tilde{f}(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t) d\Omega , \quad (3.6)$$

where we introduced the *generalised energy distribution function* $\tilde{f}(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), t)$. Thus, in the framework of spherical projections presented in Chapter 1, the expansion coefficients $f_{l,m}(\mathbf{x}, \varepsilon, t)$ correspond in fact to the expansion coefficients of the generalised energy distribution function \tilde{f} . In the following the dependence of \mathbf{k} on $(\varepsilon, \theta, \varphi)$ is implicitly assumed as well as the dependence of $Y_{l,m}$ on (θ, φ) to keep the expressions clear.

For reference, we repeat the BTE (2.22)

$$\frac{\partial f}{\partial t} + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f + \frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f = Q\{f\} . \quad (3.7)$$

and proceed with a projection onto spherical harmonics term by term.

- Term $\partial f / \partial t$: The time derivative is the simplest term to handle, since interchanging integration and differentiation immediately yields

$$\begin{aligned} \int_{\Omega} Y_{l,m} \frac{\partial f(\mathbf{x}, \mathbf{k}, t)}{\partial t} g(\varepsilon, \theta, \varphi) d\Omega &= \frac{\partial}{\partial t} \int_{\Omega} Y_{l,m} f(\mathbf{x}, \mathbf{k}, t) g(\varepsilon, \theta, \varphi) d\Omega \\ &= \frac{\partial f_{l,m}(\mathbf{x}, \varepsilon, t)}{\partial t} . \end{aligned} \quad (3.8)$$

- Term $\mathbf{v} \cdot \nabla_{\mathbf{x}} f$: Since the velocity \mathbf{v} does not show a spatial dependence, $\mathbf{v} \nabla_{\mathbf{x}} f = \nabla_{\mathbf{x}} \cdot (\mathbf{v} f)$ holds and the derivative can be pulled in front of the integral again:

$$\int_{\Omega} Y_{l,m} \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{k}, t) \tilde{g}(\varepsilon) d\Omega = \nabla_{\mathbf{x}} \cdot \int_{\Omega} Y_{l,m} \mathbf{v}(\mathbf{k}) f(\mathbf{x}, \mathbf{k}, t) g(\varepsilon, \theta, \varphi) d\Omega . \quad (3.9)$$

Since $v(\mathbf{k}) = v(\mathbf{k}(\varepsilon, \theta, \varphi))$, the group velocity cannot be pulled out of the integrand, thus we cannot find a direct relation with $f_{l,m}(\mathbf{x}, \varepsilon, t)$. Instead, the definition (2.77) suggests to define the *generalised current density*

$$\mathbf{j}_{l,m}(\mathbf{x}, \varepsilon, t) = \int_{\Omega} Y_{l,m} \mathbf{v}(\mathbf{k}) f(\mathbf{x}, \mathbf{k}, t) g(\varepsilon, \theta, \varphi) d\Omega . \quad (3.10)$$

The particle current density \mathbf{j} can then by definition directly be obtained from the zeroth order term in the expansion of the generalised current density as

$$\mathbf{j}(\mathbf{x}, t) := \frac{1}{Y_{0,0}} \int_0^\infty \mathbf{j}_{0,0}(\mathbf{x}, \varepsilon, t) d\varepsilon . \quad (3.11)$$

Thus, we arrive at

$$\int_{\Omega} Y_{l,m} \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{k}, t) g(\varepsilon, \theta, \varphi) d\Omega = \nabla_{\mathbf{x}} \cdot \mathbf{j}_{l,m}(\mathbf{x}, \varepsilon, t) . \quad (3.12)$$

- Term $\mathbf{F}/\hbar \cdot \nabla_{\mathbf{k}} f$: The transformation of the integral over the whole \mathbf{k} -space to an integral over energy requires an additional transformation of the gradient with respect to \mathbf{k} . Before the integral is transformed to equi-energy-surfaces, an integration by parts is carried out. To avoid formal difficulties with the delta distribution, we multiply the projection integral with a test function $\psi(\varepsilon)$ and integrate over energy:

$$\begin{aligned} & \int_0^\infty \psi(\varepsilon) \frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} \delta(\varepsilon - \varepsilon(\mathbf{k})) Y_{l,m} \left(\frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f(\mathbf{x}, \mathbf{k}, t) \right) d\mathbf{k} d\varepsilon \\ &= \frac{2}{(2\pi)^3 \hbar} \mathbf{F} \cdot \int_{\mathbb{R}^3} \psi(\varepsilon(\mathbf{k})) Y_{l,m} \nabla_{\mathbf{k}} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} \\ &= -\frac{2}{(2\pi)^3 \hbar} \mathbf{F} \cdot \int_{\mathbb{R}^3} \nabla_{\mathbf{k}} (\psi(\varepsilon(\mathbf{k})) Y_{l,m}) f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} \\ &= -\frac{2}{(2\pi)^3 \hbar} \mathbf{F} \cdot \int_{\mathbb{R}^3} [\psi'(\varepsilon(\mathbf{k})) \nabla_{\mathbf{k}} \varepsilon(\mathbf{k}) Y_{l,m} + \psi(\varepsilon(\mathbf{k})) \nabla_{\mathbf{k}} Y_{l,m}] f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} \end{aligned} \quad (3.13)$$

Recalling the expression for the gradient in spherical coordinates (k, θ, φ) in \mathbf{k} -space,

$$\nabla_{\mathbf{k}} Y_{l,m}(\theta, \phi) = \frac{\partial Y_{l,m}}{\partial k} \mathbf{e}_k + \frac{1}{k} \frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_\theta + \frac{1}{k \sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_\varphi , \quad (3.14)$$

where \mathbf{e}_k , \mathbf{e}_θ , \mathbf{e}_φ denote the unit vectors in k , θ and φ -direction respectively, we use the invariance of spherical harmonics in radial direction to obtain with $\mathbf{v} = \nabla_{\mathbf{k}} \varepsilon / \hbar$

$$\begin{aligned} & \int_0^\infty \psi(\varepsilon) \frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} \delta(\varepsilon - \varepsilon(\mathbf{k})) Y_{l,m} \left(\frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f(\mathbf{x}, \mathbf{k}, t) \right) d\mathbf{k} d\varepsilon \\ &= -\frac{2}{(2\pi)^3} \mathbf{F} \cdot \int_0^\infty \psi'(\varepsilon) \int_{\mathbb{R}^3} \delta(\varepsilon - \varepsilon(\mathbf{k})) Y_{l,m} \mathbf{v} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k} d\varepsilon \\ &\quad - 2\mathbf{F} \cdot \int_0^\infty \psi(\varepsilon) \int_{\Omega} \frac{1}{\hbar k} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_\theta + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_\varphi \right) f(\mathbf{x}, \mathbf{k}, t) g(\varepsilon, \theta, \varphi) d\Omega d\varepsilon \\ &= -\mathbf{F} \cdot \int_0^\infty \psi'(\varepsilon) \mathbf{j}_{l,m}(\mathbf{x}, \varepsilon, t) + \psi(\varepsilon) \mathbf{\Gamma}_{l,m} d\varepsilon \\ &= \mathbf{F} \cdot \int_0^\infty \psi(\varepsilon) \left(\frac{\mathbf{j}_{l,m}(\mathbf{x}, \varepsilon, t)}{\partial \varepsilon} - \mathbf{\Gamma}_{l,m} \right) d\varepsilon , \end{aligned}$$

where we have used the definition of the generalised current density (3.10) and set

$$\mathbf{\Gamma}_{l,m} := \frac{1}{\hbar k} \int_{\Omega} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_{\varphi} \right) \tilde{f}(\mathbf{x}, \mathbf{k}, t) d\Omega . \quad (3.15)$$

Since the test function ψ was arbitrary, we obtain

$$\frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} \delta(\varepsilon - \varepsilon(\mathbf{k})) Y_{l,m} \left(\frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f(\mathbf{x}, \mathbf{k}, t) \right) d\mathbf{k} = \mathbf{F} \cdot \left(\frac{\partial \mathbf{j}_{l,m}}{\partial \varepsilon} - \mathbf{\Gamma}_{l,m} \right) . \quad (3.16)$$

- Term $Q\{f\}$: We consider the scattering operator in low density approximation:

$$Q\{f\} = \int_B s(\mathbf{x}, \mathbf{k}', \mathbf{k}) f(\mathbf{x}, \mathbf{k}', t) - s(\mathbf{x}, \mathbf{k}, \mathbf{k}') f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k}' . \quad (3.17)$$

To allow for several different scattering processes like acoustic and optical phonon scattering, the index η is used and the scattering rate for a small volume V is written as

$$s(\mathbf{x}, \mathbf{k}, \mathbf{k}') = \frac{1}{V} \sum_{\eta} \sigma_{\eta}(\mathbf{x}, \mathbf{k}, \mathbf{k}') \delta(\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}') \pm \hbar \omega_{\eta}) , \quad (3.18)$$

where \mathbf{k} is the initial state and \mathbf{k}' is the final state. The minus sign stands for emission of energy and the plus sign for absorption. In case of multiple energy bands, summation over all energy bands has to be added to (3.18) (cf. [24]). In the following we consider a single band only, a generalisation to multiple bands mainly consists of summation over all energy bands.

The scattering integral is split into an in-scattering term

$$Q^{\text{in}}\{f\} = \frac{V}{(2\pi)^3} \int_{\mathbb{R}^3} s(\mathbf{x}, \mathbf{k}', \mathbf{k}) f(\mathbf{x}, \mathbf{k}', t) d\mathbf{k}' \quad (3.19)$$

and an out-scattering term

$$Q^{\text{out}}\{f\} = \frac{V}{(2\pi)^3} f(\mathbf{x}, \mathbf{k}, t) \int_{\mathbb{R}^3} s(\mathbf{x}, \mathbf{k}, \mathbf{k}') d\mathbf{k}' , \quad (3.20)$$

which we project onto spherical harmonics one after another. Projection of the in-scattering term (3.19) yields

$$\begin{aligned} Q_{l,m}^{\text{in}}\{f\} &= \frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} \delta(\varepsilon - \varepsilon(\mathbf{k})) Y_{l,m}(\theta(\mathbf{k}), \varphi(\mathbf{k})) Q^{\text{in}}\{f\} d\mathbf{k} \\ &= \sum_{\eta} \sum_{l'=0}^N \sum_{m'=-l'}^{l'} Q_{\eta,n,m,l',m'}^{\text{in}}(\mathbf{x}, \varepsilon) f_{l',m'}(\mathbf{x}, \varepsilon \pm \hbar \omega_{\eta}, t) , \end{aligned} \quad (3.21)$$

where

$$\begin{aligned} Q_{\eta,n,m,l',m'}^{\text{in}}(\mathbf{x}, \varepsilon) &= \int_{\Omega} Y_{l,m}(\theta, \varphi) g(\varepsilon, \theta, \varphi) \\ &\quad \times \int_{\Omega} \sigma_{\eta}(\mathbf{x}, \mathbf{k}(\varepsilon \pm \hbar \omega_{\eta}, \theta', \varphi'), \mathbf{k}(\varepsilon, \theta, \varphi)) Y_{l',m'}(\theta', \varphi') d\Omega' d\Omega . \end{aligned} \quad (3.22)$$

Similarly, the out-scattering term is evaluated as

$$Q_{l,m}^{\text{out}}\{f\} = \sum_{l'=0}^N \sum_{m'=-l'}^{l'} Q_{l,m,l',m'}^{\text{out}}(\mathbf{x}, \varepsilon) f_{l',m'}(\mathbf{x}, \varepsilon, t) , \quad (3.23)$$

where

$$Q_{l,m,l',m'}^{\text{out}}(\mathbf{x}, \varepsilon) = \int_{\Omega} Y_{l,m}(\theta, \varphi) Y_{l',m'}(\theta, \varphi) \left[\sum_{\eta} \int_{\Omega} g(\varepsilon \mp \hbar\omega_{\eta}, \theta', \varphi') \right. \\ \left. \times \sigma_{\eta}(\mathbf{x}, \mathbf{k}(\varepsilon, \theta, \varphi), \mathbf{k}(\varepsilon \mp \hbar\omega_{\eta}, \theta', \varphi')) d\Omega' \right] d\Omega . \quad (3.24)$$

A considerable simplification can be achieved if the transition rate is assumed to be *velocity randomising*, i.e. the coefficient σ_{η} in (3.18) depends only on the initial and final energy, but no longer on the angles. This allows us to rewrite (3.22) as follows:

$$Q_{\eta,n,m,l',m'}^{\text{in}}(\mathbf{x}, \varepsilon) = \sigma_{\eta}(\mathbf{x}, \varepsilon \pm \hbar\omega_{\eta}, \varepsilon) \times \int_{\Omega} Y_{l,m} g(\varepsilon, \theta, \varphi) d\Omega \times \int_{\Omega} Y_{l',m'} d\Omega' \\ = \frac{1}{Y_{0,0}} \sigma_{\eta}(\mathbf{x}, \varepsilon \pm \hbar\omega_{\eta}, \varepsilon) g_{l,m}(\varepsilon) \delta_{0,l'} \delta_{0,m'} , \quad (3.25)$$

where $g_{l,m}(\varepsilon)$ denotes the projection of the generalised density of states onto the spherical harmonic $Y_{l,m}$. In case of spherical bands, only $f_{0,0}$ is coupled into the balance equation for $l = m = 0$ since then $g_{l,m}(\varepsilon) = 0$ for $(l, m) \neq (0, 0)$. The out-scattering term can under the assumption of velocity-randomisation be simplified to

$$Q_{l,m,l',m'}^{\text{out}}(\mathbf{x}, \varepsilon) = \sum_{\eta} \sigma_{\eta}(\mathbf{x}, \varepsilon, \varepsilon \mp \hbar\omega_{\eta}) \int_{\Omega} Y_{l,m} Y_{l',m'} d\Omega \times \int_{\Omega} \tilde{g}(\varepsilon \mp \hbar\omega_{\eta}) d\Omega' \\ = \frac{1}{Y_{0,0}} \sum_{\eta} \sigma_{\eta}(\mathbf{x}, \varepsilon, \varepsilon \mp \hbar\omega_{\eta}) g_{0,0}(\varepsilon \mp \hbar\omega_{\eta}) \delta_{l,l'} \delta_{m,m'} . \quad (3.26)$$

Thus, the out-scattering term is proportional to $f_{l,m}$.

Summing up, the full projected scattering operator using velocity randomisation (VR) is thus given by

$$Q_{l,m}^{\text{VR}}\{f\} = \frac{1}{Y_{0,0}} \left[g_{l,m}(\varepsilon) \sum_{\eta} \sigma_{\eta}(\mathbf{x}, \varepsilon \pm \hbar\omega_{\eta}, \varepsilon) f_{0,0}(\mathbf{x}, \varepsilon \pm \hbar\omega_{\eta}, t) \right. \\ \left. - f_{l,m}(\mathbf{x}, \varepsilon, t) \sum_{\eta} \sigma_{\eta}(\mathbf{x}, \varepsilon, \varepsilon \mp \hbar\omega_{\eta}) g_{0,0}(\varepsilon \mp \hbar\omega_{\eta}) \right] . \quad (3.27)$$

Even though the scattering operator after projection is not an integral operator anymore, deviated arguments on the right hand side show up whenever inelastic collisions with $\hbar\omega_{\eta} \neq 0$ are considered. This complicates analytical investigations even more, but as we will see in the next chapter, a discretisation can still be obtained in a rather straight-forward fashion.

Collecting all individual terms, we obtain for the SHE of the BTE the system of coupled balance equations under the assumption of velocity randomisation

$$\begin{aligned} \frac{\partial f_{l,m}}{\partial t} + \nabla_{\mathbf{x}} \cdot \mathbf{j}_{l,m}(\mathbf{x}, \varepsilon, t) + \mathbf{F} \cdot \left(\frac{\partial \mathbf{j}_{l,m}}{\partial \varepsilon} - \mathbf{\Gamma}_{l,m} \right) = \\ = \frac{1}{Y_{0,0}} \left[g_{l,m}(\varepsilon) \sum_{\eta} \sigma_{\eta}(\mathbf{x}, \varepsilon \pm \hbar\omega_{\eta}, \varepsilon) f_{0,0}(\mathbf{x}, \varepsilon \pm \hbar\omega_{\eta}, t) \right. \\ \left. - f_{l,m}(\mathbf{x}, \varepsilon, t) \sum_{\eta} \sigma_{\eta}(\mathbf{x}, \varepsilon, \varepsilon \mp \hbar\omega_{\eta}) g_{0,0}(\varepsilon \mp \hbar\omega_{\eta}) \right]. \end{aligned} \quad (3.28)$$

Finally, we can collect the derivatives on the left hand side to obtain a single divergence in the $(\mathbf{x}, \varepsilon)$ -space:

$$\tilde{\nabla} = \begin{pmatrix} \nabla_{\mathbf{x}} \\ \frac{\partial}{\partial \varepsilon} \end{pmatrix}, \quad \tilde{\mathbf{j}}_{l,m} = \begin{pmatrix} \mathbf{j}_{l,m} \\ \mathbf{F} \cdot \mathbf{j}_{l,m} \end{pmatrix} \quad (3.29)$$

and rewrite (3.28) as

$$\frac{\partial f_{l,m}}{\partial t} + \tilde{\nabla} \cdot \tilde{\mathbf{j}}_{l,m}(\mathbf{x}, \varepsilon, t) - \mathbf{F} \cdot \mathbf{\Gamma}_{l,m} = Q_{l,m}^{\text{VR}}\{f\}. \quad (3.30)$$

In this divergence form, the derivation of a discretisation scheme is simplified. However, it has been observed that a direct discretisation of (3.30) suffers from similar numerical instabilities similar to those observed at the drift-diffusion model [43]. The remainder of this chapter is thus devoted to the construction of a suitable (continuous) stabilisation.

3.2 Parity Splitting

Average quantities defined as moments of the distribution function as described in Section 2.4 suggest to consider even and odd parts of the distribution function with respect to the wave vector \mathbf{k} :

$$f^{\text{even}}(\mathbf{x}, \mathbf{k}, t) = \frac{f(\mathbf{x}, \mathbf{k}, t) + f(\mathbf{x}, -\mathbf{k}, t)}{2}, \quad (3.31)$$

$$f^{\text{odd}}(\mathbf{x}, \mathbf{k}, t) = \frac{f(\mathbf{x}, \mathbf{k}, t) - f(\mathbf{x}, -\mathbf{k}, t)}{2}. \quad (3.32)$$

The even part f^{even} then carries information about scalar valued macroscopic quantities such as the carrier density or average energy¹ (the zeroth and second moment respectively), while the odd part determines vector valued macroscopic quantities such as the current density.

With the shorthand notation

$$L\{f\} := \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \frac{1}{\hbar} \mathbf{F} \nabla_{\mathbf{k}} f \quad (3.33)$$

for the *free streaming operator* and the assumption that the velocity \mathbf{v} is an even function of the wave vector \mathbf{k} (which can always be assumed for reasons of symmetry of the physical processes), we observe the following:

¹under the assumption of spherical bands

Lemma 6. Let $f(\mathbf{x}, \mathbf{k}, t) \in C^\infty := C^\infty(\mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R})$ and C^{even} and C^{odd} be the subspaces of even and odd functions of C^∞ with respect to \mathbf{k} . Then

$$f \in C^{\text{even}} \implies L\{f\} \in C^{\text{odd}}, \quad (3.34)$$

$$f \in C^{\text{odd}} \implies L\{f\} \in C^{\text{even}}. \quad (3.35)$$

Proof. First we note that the gradient of an even function is an odd function and vice versa.

Now, let $f^{\text{even}} \in C^{\text{even}}$, then

$$L\{f^{\text{even}}\} = \underbrace{\mathbf{v}}_{\text{odd in } \mathbf{k}} \cdot \underbrace{\nabla_{\mathbf{x}} f^{\text{even}}}_{\text{even in } \mathbf{k}} + \frac{1}{\hbar} \mathbf{F} \cdot \underbrace{\nabla_{\mathbf{k}} f^{\text{even}}}_{\text{odd in } \mathbf{k}} \in C^{\text{odd}}. \quad (3.36)$$

Similarly, for $f^{\text{odd}} \in C^{\text{even}}$ we obtain

$$L\{f^{\text{odd}}\} = \underbrace{\mathbf{v}}_{\text{odd in } \mathbf{k}} \cdot \underbrace{\nabla_{\mathbf{x}} f^{\text{odd}}}_{\text{odd in } \mathbf{k}} + \frac{1}{\hbar} \mathbf{F} \cdot \underbrace{\nabla_{\mathbf{k}} f^{\text{odd}}}_{\text{even in } \mathbf{k}} \in C^{\text{even}}. \quad (3.37)$$

□

Similar considerations for the collision operator $Q\{f\}$ in the low density approximation (2.41) need assumptions for the parity of the scattering cross section $\sigma(\mathbf{x}, \mathbf{k}, \mathbf{k}')$. Again, we may assume from symmetry of the physical processes that σ is even in its wave vectors, which allows to extract parity information from the collision operator:

Lemma 7. For even scattering cross sections $\sigma(\mathbf{x}, \mathbf{k}, \mathbf{k}')$ with respect to the wave vectors in the low density approximation (2.41) of the collision operator Q , parity with respect to the wave vector \mathbf{k} is conserved, i.e.

$$f \in C^{\text{even}} \implies Q\{f\} \in C^{\text{even}}, \quad (3.38)$$

$$f \in C^{\text{odd}} \implies Q\{f\} \in C^{\text{odd}}. \quad (3.39)$$

Proof. This statement follows immediately from the low density approximation (2.41). □

Due to the parity properties of the free streaming operator, the BTE for the distribution function $f = f^{\text{even}} + f^{\text{odd}}$ can be written for the even and odd part separately:

$$\frac{\partial f^{\text{even}}}{\partial t} + L\{f^{\text{odd}}\} = Q^{\text{even}}\{f\}, \quad (3.40)$$

$$\frac{\partial f^{\text{odd}}}{\partial t} + L\{f^{\text{even}}\} = Q^{\text{odd}}\{f\}, \quad (3.41)$$

where Q^{even} and Q^{odd} denote the even and odd part of a general collision operator $Q\{f\}$. If we additionally assume a low density approximation with even collision cross sections as in Lemma 7, we can write

$$\frac{\partial f^{\text{even}}}{\partial t} + L\{f^{\text{odd}}\} = Q\{f^{\text{even}}\}, \quad (3.42)$$

$$\frac{\partial f^{\text{odd}}}{\partial t} + L\{f^{\text{even}}\} = Q\{f^{\text{odd}}\}. \quad (3.43)$$

Thus, a natural requirement on expansion methods for the solution of the BTE is that the parity coupling due to the free streaming operator is preserved. More precisely, we expect that the drift term couples even spherical harmonics $Y_{l,m}$ to odd ones, but does not couple harmonics of the same parity.

Since the (generalised) density of states g is an even function of the wave vector, the distribution function f is of the same parity as the generalised energy distribution function \tilde{f} . Moreover, from Lemma 2 we deduce that the even and the odd part (with respect to the wave vector) of \tilde{f} are given as

$$\tilde{f}^{\text{even}}(\mathbf{x}, \varepsilon, \theta, \varphi, t) = \sum_{l=0}^{\infty} \sum_{m=-2i}^{2i} f_{2l,m}(\mathbf{x}, \varepsilon, t) Y_{2l,m}(\theta, \varphi) , \quad (3.44)$$

$$\tilde{f}^{\text{odd}}(\mathbf{x}, \varepsilon, \theta, \varphi, t) = \sum_{l=0}^{\infty} \sum_{m=-2i-1}^{2i+1} f_{2l+1,m}(\mathbf{x}, \varepsilon, t) Y_{2l+1,m}(\theta, \varphi) . \quad (3.45)$$

The drift operator in the SHE equation (3.30) consists of the two terms $\mathbf{j}_{l,m}$ and $\mathbf{\Gamma}_{l,m}$, which should consequently couple spherical harmonics of different parity only. Inserting the expansion of \tilde{f} into $\mathbf{j}_{l,m}$ yields

$$\begin{aligned} \mathbf{j}_{l,m}(\mathbf{x}, \varepsilon, t) &= \int_{\Omega} Y_{l,m} \mathbf{v}(\mathbf{k}) \tilde{f}(\mathbf{x}, \varepsilon, \theta, \varphi, t) d\Omega \\ &= \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} f_{l',m'}(\mathbf{x}, \varepsilon, t) \underbrace{\int_{\Omega} Y_{l,m} \mathbf{v}(\mathbf{k}) Y_{l',m'} d\Omega}_{:=\mathbf{v}_{l,m,l',m'}} . \end{aligned} \quad (3.46)$$

Since the velocity $\mathbf{v}(\mathbf{k}) = \nabla_{\mathbf{k}} \varepsilon(\mathbf{k}) / \hbar$ is an odd function of \mathbf{k} , we obtain for reasons of parity with Lemma 2 that

$$\mathbf{v}_{2l,m,2l',m'} = \mathbf{0} \quad \text{and} \quad \mathbf{v}_{2l+1,m,2l'+1,m'} = \mathbf{0} . \quad (3.47)$$

On the other hand, for the term $\mathbf{\Gamma}_{l,m}$ we obtain

$$\begin{aligned} \mathbf{\Gamma}_{l,m} &= \int_{\Omega} \frac{1}{\hbar k} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_{\varphi} \right) \tilde{f}(\mathbf{x}, \mathbf{k}, t) d\Omega \\ &= \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} f_{l',m'}(\mathbf{x}, \varepsilon, t) \underbrace{\int_{\Omega} \frac{1}{\hbar k} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_{\varphi} \right) Y_{l',m'} d\Omega}_{=:\mathbf{\Gamma}_{l,m,l',m'}} . \end{aligned} \quad (3.48)$$

For parity considerations, the unit vectors must not be left unconsidered:

$$\mathbf{e}_{\theta}(\pi - \theta, \pi + \varphi) = \begin{pmatrix} \cos(\pi - \theta) \cos(\pi + \varphi) \\ \cos(\pi - \theta) \sin(\pi + \varphi) \\ -\sin(\pi - \theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) \cos(\varphi) \\ \cos(\theta) \sin(\varphi) \\ -\sin(\theta) \end{pmatrix} = \mathbf{e}_{\theta}(\theta, \varphi) , \quad (3.49)$$

$$\mathbf{e}_{\varphi}(\pi - \theta, \pi + \varphi) = \begin{pmatrix} -\sin(\pi + \varphi) \\ \cos(\pi + \varphi) \\ 0 \end{pmatrix} = -\mathbf{e}_{\varphi}(\theta, \varphi) . \quad (3.50)$$

Since $k = |\mathbf{k}|$ is an even function, it does not enter into further parity considerations for (3.48). From Lemma 3 together with the parity of the unit vectors (3.49) and (3.50) we see that the term

$$\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_\theta + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_\varphi$$

is of opposite parity than $Y_{l,m}$, thus

$$\mathbf{\Gamma}_{2l,m,2l',m'} = \mathbf{0} \quad \text{and} \quad \mathbf{\Gamma}_{2l+1,m,2l'+1,m'} = \mathbf{0} . \quad (3.51)$$

We therefore conclude that the odd/even coupling of the free streaming operator is preserved under SHE.

3.3 Entropy Considerations

The hyperbolic nature of the BTE leads to numerical difficulties that have been reported and investigated in several publications [9, 34–37, 43]. In this section a stabilisation scheme based on entropies pioneered by Ringhofer [34, 35, 37] is discussed. For the sake of conciseness, the opposite sign of the collision operator in the works of Ringhofer has been adjusted to meet the notational conventions used in this thesis.

Even though entropy is a concept tightly related with physics, we start with a mathematical definition:

Definition 9. *An entropy is a function $H(f, \mathbf{x}, \mathbf{k})$ satisfying the following properties:*

(i) *The entropy is dissipated by the collision operator:*

$$\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} Q\{f\} d\mathbf{k} d\mathbf{x} \leq 0 . \quad (3.52)$$

(ii) *The entropy is preserved by the free streaming operator, so*

$$\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} \left[\mathbf{v} \cdot \nabla_{\mathbf{x}} f + \frac{1}{\hbar} \mathbf{F}(\mathbf{x}) \cdot \nabla_{\mathbf{k}} f \right] d\mathbf{k} d\mathbf{x} = 0 \quad (3.53)$$

holds.

(iii) *For fixed \mathbf{x} and \mathbf{k} the function $H(f, \mathbf{x}, \mathbf{k})$ is a strictly convex function of the variable f , i.e.*

$$H(tf + (1-t)g, \mathbf{x}, \mathbf{k}) < tH(f, \mathbf{x}, \mathbf{k}) + (1-t)H(g, \mathbf{x}, \mathbf{k}) \quad \forall t \in (0, 1), \quad f \neq g , \quad (3.54)$$

and the entropy functional

$$\eta(f) = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} H(f, \mathbf{x}, \mathbf{k}) d\mathbf{x} d\mathbf{k} \quad (3.55)$$

is a convex functional of f .

We can directly deduce the following property of the entropy functional:

Lemma 8. *The functional η is non-increasing in time.*

Proof. We multiply the BTE (2.22) with $\partial H(f, \mathbf{x}, \mathbf{k})/\partial f$ and integrate:

$$\begin{aligned} & \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} \frac{\partial f}{\partial t} d\mathbf{k} d\mathbf{x} \\ & + \underbrace{\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f + \frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f d\mathbf{k} d\mathbf{x}}_{=0} = \underbrace{\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} Q\{f\} d\mathbf{k} d\mathbf{x}}_{\leq 0} . \end{aligned} \quad (3.56)$$

Consequently, we obtain

$$0 \geq \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} \frac{\partial f}{\partial t} d\mathbf{k} d\mathbf{x} = \frac{d}{dt} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} H(f, \mathbf{x}, \mathbf{k}) d\mathbf{k} d\mathbf{x} = \frac{d}{dt} \eta(f) , \quad (3.57)$$

hence the functional η is non-increasing in time. \square

An entropy based Galerkin method can now be constructed as follows: We first introduce the inverse of $\partial H(f, \mathbf{x}, \mathbf{k})/\partial f$ with respect to the variable f by

$$\tilde{f} := \frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} \iff f = \mu(\tilde{f}, \mathbf{x}, \mathbf{k}) , \quad (3.58)$$

which exists since $\partial H(f, \mathbf{x}, \mathbf{k})/\partial f$ is a strictly monotone function in f . The Galerkin procedure is now applied to the entropy variable \tilde{f} , which is expanded in a set of basis functions $\{\varphi_m(\mathbf{k}), m = 0, 1, \dots, M\}$ as

$$\tilde{f}(\mathbf{x}, \mathbf{k}, t) \simeq \sum_{m=0}^M \tilde{f}_m(\mathbf{x}, t) \varphi_m(\mathbf{k}) , \quad (3.59)$$

and the distribution function f can be obtained from this expansion of \tilde{f} as

$$f(\mathbf{x}, \mathbf{k}, t) = \mu\left(\sum_{m=0}^M \tilde{f}_m(\mathbf{x}, t) \varphi_m(\mathbf{k}), \mathbf{x}, t\right) . \quad (3.60)$$

Writing the BTE in the entropy variable $\tilde{f}(\mathbf{x}, \mathbf{k}, t)$, a projection onto each of the basis functions yields a system of partial differential equations of first order (this system is given explicitly e.g. in [37]). By virtue of construction, such a system automatically satisfies

$$\frac{d}{dt} \tilde{\eta}(\mathbf{F}) \leq 0 \quad (3.61)$$

where \mathbf{F} is the coefficient vector of the system and $\tilde{\eta}$ is the discrete analogue of the entropy functional (3.55). The actual form of the resulting system depends on two choices:

- The set of basis functions.
- The entropy function $H(f, \mathbf{x}, \mathbf{k})$.

The basis used in this thesis was a-priori already chosen to be the set of spherical harmonics to account for certain symmetries of the dispersion relation. Thus, we will further investigate possible entropy functions.

In classical physics the entropy is given as the natural logarithm:

$$\frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} = \ln f \iff H(f, \mathbf{x}, \mathbf{k}) = f(\ln f - 1) . \quad (3.62)$$

However, since the collision operator is often simplified in one way or another and we are additionally considering particles driven by an external field, there is a much larger degree of freedom in the choice of the entropy.

In the following formal derivation, we assume a low density approximation (2.41) of the collision operator $Q\{f\}$ with symmetric collision cross-sections. This assumption allows the following statement [18]:

Lemma 9. *A sufficient condition on the entropy $H(f, \mathbf{x}, \mathbf{k})$ is that its derivative is a non-decreasing function of $f(\mathbf{x}, \mathbf{k}, t) \exp(\varepsilon(\mathbf{k})/(k_B T))$, i.e.*

$$\frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} = h_1\left(f \exp\left(\frac{\varepsilon(\mathbf{k})}{k_B T}\right), \mathbf{x}\right) \quad \text{with} \quad \frac{\partial h_1(u, \mathbf{x})}{\partial u} \geq 0 , \quad (3.63)$$

the dissipation property (3.52) is fulfilled.

Proof. We consider the dissipation property (3.52). For brevity, we omit function arguments and write H , f if evaluated at \mathbf{k} and H' , f' if evaluated at \mathbf{k}' . With shorthand notation $\mathcal{E} := \exp(\varepsilon(\mathbf{k})/(k_B T))$ and $\mathcal{E}' := \exp(\varepsilon(\mathbf{k}')/(k_B T))$ we find

$$\begin{aligned} \int_{\mathbb{R}^3} \frac{\partial H}{\partial f} Q\{f\} d\mathbf{k} &= \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \sigma \frac{\partial H}{\partial f} [\mathcal{E} f - \mathcal{E}' f'] d\mathbf{k}' d\mathbf{k} \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \sigma \frac{\partial H}{\partial f} [\mathcal{E} f - \mathcal{E}' f'] d\mathbf{k}' d\mathbf{k} \\ &\quad + \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \sigma \frac{\partial H}{\partial f} [\mathcal{E} f - \mathcal{E}' f'] d\mathbf{k}' d\mathbf{k} \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \sigma \frac{\partial H}{\partial f} [\mathcal{E} f - \mathcal{E}' f'] d\mathbf{k}' d\mathbf{k} \\ &\quad - \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \sigma \frac{\partial H'}{\partial f} [\mathcal{E} f - \mathcal{E}' f'] d\mathbf{k}' d\mathbf{k} \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \sigma \left[\frac{\partial H}{\partial f} - \frac{\partial H'}{\partial f} \right] [\mathcal{E} f - \mathcal{E}' f'] d\mathbf{k}' d\mathbf{k} , \end{aligned}$$

where we used that the collision cross-section is symmetric with respect to \mathbf{k} and \mathbf{k}' and that interchanging the roles of \mathbf{k} and \mathbf{k}' in the double integral changes the sign.

Since the collision cross-section σ is nonnegative, any entropy H whose derivative is a non-decreasing function of $\mathcal{E} f$, i.e.

$$\frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} = h_1\left(f \exp\left(\frac{\varepsilon(\mathbf{k})}{k_B T}\right), \mathbf{x}\right) \quad \text{with} \quad \frac{\partial h_1(u, \mathbf{x})}{\partial u} \geq 0 , \quad (3.64)$$

satisfies the dissipation property, because the term $[\partial H/\partial f - \partial H'/\partial f][\mathcal{E} f - \mathcal{E}' f']$ in the integrand is then always nonnegative. \square

Before we proceed with a closer look at the entropy conservation (3.53), some preparatory work is required:

Lemma 10. *Let $H(f, \mathbf{x}, \mathbf{k})$ be an entropy, where f satisfies the BTE (2.22). Provided that $H(f, \mathbf{x}, \mathbf{k})$ goes sufficiently fast to zero as $|\mathbf{x}|, |\mathbf{k}| \rightarrow \infty$, there holds*

$$\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \nabla_{\mathbf{x}} \cdot (\mathbf{v}H) + \frac{1}{\hbar} \nabla_{\mathbf{k}} \cdot (\mathbf{F}H) d\mathbf{k} d\mathbf{x} = 0 . \quad (3.65)$$

Proof. We select a cube C_l in $\mathbb{R}^3 \times \mathbb{R}^3$ with edge length l centered at the origin and consider the limit $l \rightarrow \infty$. With Gauss' Law we obtain

$$\int_{C_l} \nabla_{\mathbf{x}} \cdot (\mathbf{v}H) + \nabla_{\mathbf{k}} \cdot (\mathbf{F}H) dV = \int_{\partial C_l} \left(\frac{\mathbf{v}}{\frac{1}{\hbar} \mathbf{F}} \right) H d\mathbf{A} , \quad (3.66)$$

where \mathbf{A} is the area element perpendicular to the surface of the cube. In the limit $l \rightarrow \infty$ we see that (3.66) reduces to (3.65), since by assumption $H(f, \mathbf{x}, \mathbf{k})$ decays sufficiently fast to zero. \square

Note that because of the definition of the entropy functional $\eta(f)$ in (3.55), $H(f, \mathbf{x}, \mathbf{k})$ is required to go to zero as \mathbf{x} and \mathbf{k} grow in magnitude, otherwise the entropy functional would be infinity and thus be useless. However, $H(f, \mathbf{x}, \mathbf{k})$ is additionally required to dominate \mathbf{v} and \mathbf{F} in (3.66), where at least the former is known to be unbounded as $|\mathbf{k}|$ goes to infinity.

In order to obtain entropy conservation by the free streaming operator given in (3.53), the integrand is written as

$$\begin{aligned} \frac{\partial H}{\partial f} \left[\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} f + \frac{1}{\hbar} \mathbf{F}(\mathbf{x}) \cdot \nabla_{\mathbf{k}} f \right] &= \left[\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{x}} H + \frac{1}{\hbar} \mathbf{F}(\mathbf{x}) \cdot \nabla_{\mathbf{k}} H \right] \\ &\quad - \left[\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{2}} H + \frac{1}{\hbar} \mathbf{F}(\mathbf{x}) \cdot \nabla_{\mathbf{3}} H \right] , \end{aligned} \quad (3.67)$$

where $\nabla_{\mathbf{2}}$ and $\nabla_{\mathbf{3}}$ denote partial derivatives with respect to the second and third argument of $H(f, \mathbf{x}, \mathbf{k})$ respectively. The integral over the phase space over the term in the first bracket on the right hand side vanishes due to Lemma 10. Thus, if

$$\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{2}} H + \frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{3}} H = \frac{1}{\hbar} \nabla_{\mathbf{k}} \varepsilon(\mathbf{k}) \cdot \nabla_{\mathbf{2}} H - \frac{q}{\hbar} \nabla_{\mathbf{x}} \psi(\mathbf{x}) \cdot \nabla_{\mathbf{3}} H = 0 , \quad (3.68)$$

entropy conservation is fulfilled. The general solution of this first order partial differential equation is given by all functions of the total energy $\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})$. Together with the first requirement in Lemma 9 we therefore conclude that

$$\frac{\partial H(f, \mathbf{x}, \mathbf{k})}{\partial f} = h \left(f \exp \left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T} \right) \right) \text{ with } h'(u) > 0 . \quad (3.69)$$

The physical entropy, the logarithm, now corresponds to the choice

$$h(u) = \ln u , \quad H(f, \mathbf{x}, \mathbf{k}) = f(\ln f - 1 + \varepsilon(\mathbf{k}) + q\psi(\mathbf{x})) \quad (3.70)$$

and the connection between f and its entropy variable \tilde{f} is given as

$$f = \mu(g, \mathbf{x}, \mathbf{k}) = \exp \left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T} \right) . \quad (3.71)$$

However, the exponential dependence of f on \tilde{f} causes problems both from the numerical point of view and the analytic point of view [26,37]. For numerical applications, the simplest possible entropy is chosen:

$$h(u) = u, \quad H(f, \mathbf{x}, \mathbf{k}) = \frac{f^2}{2} \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right), \quad (3.72)$$

so that the connection between f and its entropy variable \tilde{f} is now of the form

$$f = \mu(g, \mathbf{x}, \mathbf{k}) = \tilde{f} \exp\left(-\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right). \quad (3.73)$$

The crucial point is that f is now linearly related to \tilde{f} , which means that a Galerkin scheme for the entropy variable leads to a system of linear equations for a given external potential and a linear scattering operator. Additionally, the entropy functional $\eta(f)$ collapses to an L^2 estimate with weight function $\exp((\varepsilon(\mathbf{k}) + q\psi(\mathbf{x}))/k_B T)$.

Chapter 4

Discretisation

Up to now we have dealt with the continuous form of the Boltzmann transport equation (BTE) and derived a spherical harmonics expansion (SHE) in an analytic way. For an implementation in a computer, the continuous formulation has to be transferred to a discrete representation, which is often more complicated than the formulation of a suitable mathematical representation. In this chapter the transition from the continuous formulation as given in the preceding chapter to a full discretisation is described, where we start with a rather wide-spread numerical method, the *finite volume method*. After that, the Scharfetter-Gummel discretisation is introduced as a stabilisation technique to circumvent instabilities that otherwise arise in the drift-diffusion model. After that, we turn towards a suitable and stable discretisation of the SHE of the BTE.

4.1 The Finite Volume Method

Besides differencing schemes and the *finite element method* (FEM), the *finite volume method* (FVM, often also referred to as *box integration method* or *spectral volume method*) has become a wide-spread discretisation scheme, especially in the fields of semiconductor device simulation and fluid dynamics. The main advantage of the FVM is its local conservation property, thus the method is particularly suited for the discretisation of conservation equations such as the drift-diffusion equations. In this section we briefly give the basic concepts and summarise the main results. Several remarks on the differences to the typically more familiar FEM are given. More details on the mathematical background can for example be found in the textbook of Bey [5].

We consider the problem

$$\nabla \cdot (-\mathbf{D}\nabla u + \mathbf{b}u) = f \quad \text{in } \Xi \subset \mathbb{R}^n, \quad (4.1)$$

$$u = 0 \quad \text{on } \partial\Xi \quad (4.2)$$

which we will refer to as *strong formulation*. We assume that the diffusion matrix \mathbf{D} is symmetric and positive definite and that each entry of \mathbf{D} and the vector \mathbf{b} lies in $L^\infty(\Xi)$ as well as $\nabla \cdot \mathbf{b} \in L^\infty(\Xi)$. The right hand side f is assumed to lie in $H^{-1}(\Xi)$ to guarantee well-posedness of the problem. The weak formulation of this problem is found in the typical way by multiplication with a test function $v \in H_0^1(\Xi)$, integration over the whole domain

and integration by parts: We then seek $u \in H_0^1(\Xi)$ such that

$$\int_{\Xi} \nabla v \cdot \mathbf{D} \nabla u \, d\mathbf{x} - \int_{\Xi} \nabla v \cdot \mathbf{b} u \, d\mathbf{x} = \int_{\Xi} f v \, d\mathbf{x} \quad (4.3)$$

for all $v \in H_0^1(\Xi)$.

While the FEM directly applies a Galerkin method to the weak formulation (4.3), the FVM includes a decomposition of the problem domain by a tessellation $\mathcal{B} = \{B_1, \dots, B_m\}$ before partial integration is carried out, hence the left hand side of the strong formulation (4.1) can be cast into a weak form as follows:

$$\begin{aligned} \int_{\Xi} v \nabla \cdot (-\mathbf{D} \nabla u + \mathbf{b} u) \, d\mathbf{x} &= \sum_{B \in \mathcal{B}} \int_B v \nabla \cdot (-\mathbf{D} \nabla u + \mathbf{b} u) \, d\mathbf{x} \\ &= \sum_{B \in \mathcal{B}} \left[\int_B \nabla v \cdot (\mathbf{D} \nabla u - \mathbf{b} u) \, d\mathbf{x} - \int_{\partial B} v (\mathbf{D} \nabla u - \mathbf{b} u) \cdot d\mathbf{n} \right], \end{aligned} \quad (4.4)$$

where \mathbf{n} denotes the unit, outward-pointing normal vector. We thus arrive at a generalised weak formulation: Find u in a suitable solution space such that

$$\sum_{B \in \mathcal{B}} \left[\int_B \nabla v \cdot (\mathbf{D} \nabla u - \mathbf{b} u) \, d\mathbf{x} - \int_{\partial B} v (\mathbf{D} \nabla u - \mathbf{b} u) \cdot d\mathbf{n} \right] = \int_{\Xi} f v \, d\mathbf{x} \quad (4.5)$$

for all $v \in H_0^1(\mathcal{B})$, where $H_0^1(\mathcal{B})$ is the space of all functions of $L^2(\Xi)$ that are piecewise H^1 -regular with respect to the tessellation \mathcal{B} and vanish on $\partial\Xi$ in the sense of a trace. For a discussion of solution spaces we refer to the literature [5], for simplicity we assume $u \in H^2(\Xi) \cap H_0^1(\Xi)$. For the discretisation we choose v to be piecewise constant, so that we obtain the weak form

$$a(u, v) := - \sum_{B \in \mathcal{B}} \int_{\partial B} v (\mathbf{D} \nabla u - \mathbf{b} u) \cdot d\mathbf{n} = \int_{\Xi} f v \, d\mathbf{x} =: L(v). \quad (4.6)$$

The tessellation \mathcal{B} of the problem domain Ξ is typically based on a triangulation \mathcal{T} of Ξ . We mention two wide-spread methods (cf. Fig. 4.1):

- A commonly used technique in two dimensions is to use so-called *Voronoi diagrams* (aka. *circumcenter method*) [31], where the boxes are convex sets associated with a given set of disjoint points $P = \{\mathbf{p}_1, \dots, \mathbf{p}_m\}$. Each box $B_i \in \mathcal{B}$ is then given as the set of points that are closer to \mathbf{p}_i than to any other point of P , i.e.

$$B_i = \{ \mathbf{x} \in \Xi \mid \|\mathbf{x} - \mathbf{p}_i\| \leq \|\mathbf{x} - \mathbf{p}_j\| \quad \forall j \neq i, j \in \{1, \dots, m\} \} \quad (4.7)$$

An interesting fact is that each Voronoi diagram in two dimensions is the dual of a triangulation where the vertices are exactly the points of P . The converse is not true, since if any of the triangular angles exceeds $\pi/2$, the circumcenter is located outside the triangle (and hence may also be located outside the problem domain), so the resulting Voronoi diagram is not the dual of the triangulation anymore. Conversely, triangulations that are duals of Voronoi diagrams have inner angles smaller or equal to $\pi/2$ and are called *Delaunay triangulations*. The advantage of Voronoi diagrams

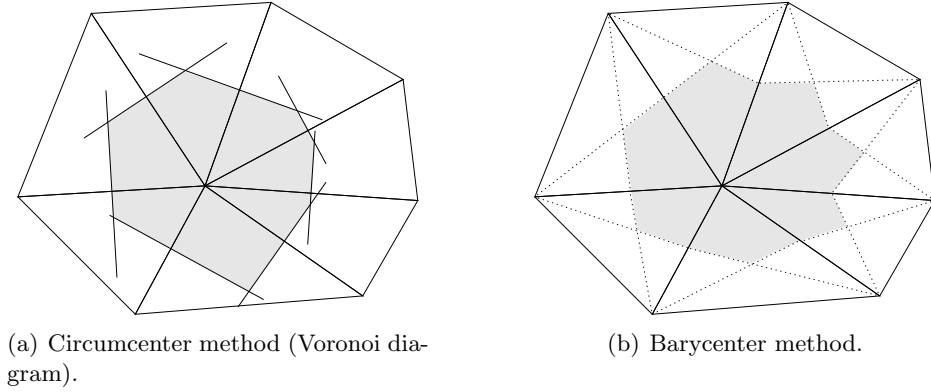


Figure 4.1: Two methods for the construction of a dual box tessellation from a triangulation.

is that the interfaces between boxes are perpendicular to the connections between the vertices centered in the boxes. However, the requirement of underlying Delaunay triangulations results in additional effort needed for mesh generation and adaptive refinement. Additionally, generalisations to higher dimensions become rather hard especially from the programmer's point of view.

- In the *barycenter method* [5] the barycenters of the triangles are connected with the midpoint of the edges to obtain the dual box tessellation (cf. Fig. 4.1(b)). An advantage is that the method can be, unlike others, algorithmically generalised to higher dimensions with reasonable effort and without further requirements on the mesh. However, in comparison to the circumcenter method the number of facets of each box is larger and the boxes are not necessarily convex.

Let us assume that the problem space Ξ is triangulated by a set \mathcal{T} and that the tessellation \mathcal{B} is a dual grid of \mathcal{T} constructed by one of the two methods described above. To proceed with a discretisation, we introduce appropriate nomenclature:

$$\mathcal{P}_s(B) := \{v \in L^2(B) \mid v \text{ is a polynomial of degree } s \text{ on } B\} \quad (4.8)$$

$$\mathcal{P}_s(\mathcal{B}) := \{v \in L^2(\Xi) \mid v|_B \in \mathcal{P}_s(B) \quad \forall B \in \mathcal{B}\} \quad (4.9)$$

Analogous definitions apply for a triangulation \mathcal{T} .

In the following we consider the cases $s = 0$ and $s = 1$ only. A basis $X(\mathcal{B})$ of $\mathcal{P}_0(\mathcal{B})$ is given by the characteristic functions for each box:

$$\chi_j \in P_0(\mathcal{B}), \quad \chi_j|_{B_i} = \delta_{ij} \quad 1 \leq i, j \leq m. \quad (4.10)$$

In a similar manner a basis for piecewise linear functions on \mathcal{T} with the set of vertices $P = \{\mathbf{p}_1, \dots, \mathbf{p}_m\}$ is given by

$$\varphi_j \in P_1(\mathcal{T}), \quad \varphi_j(\mathbf{p}_i) = \delta_{ij} \quad 1 \leq i, j \leq m. \quad (4.11)$$

Note that this basis is the most commonly used basis in FEM.

The finite volume discretisation then is to apply a non-conforming ansatz for u and v in (4.6):

$$u_h = \sum_{i=1}^m u_i \varphi_i, \quad \varphi_i \in \Phi(\mathcal{T}) \quad \forall i = \{1, \dots, m\}, \quad (4.12)$$

$$v_h = \sum_{i=1}^m v_i \chi_i, \quad \chi_i \in X(\mathcal{B}) \quad \forall i = \{1, \dots, m\}. \quad (4.13)$$

The discrete formulation thus is to find $u_h \in \mathcal{P}_1(\mathcal{T})$ such that

$$a(u_h, v_h) = L(v_h) \quad \forall v_h \in \mathcal{P}_0(\mathcal{B}), \quad (4.14)$$

which can – by the use of linearity of the weak formulation and linear independence of the basis functions – be written in matrix-vector form as

$$\mathbf{A}_h \mathbf{u}_h = \mathbf{l}_h, \quad (4.15)$$

where the (i, j) -th entry of the system matrix \mathbf{A}_h is given by $a(\varphi_j, \chi_i)$ and the i -th entry of the load vector \mathbf{l}_h by $L(\chi_i)$.

It can be shown (see e.g. [5]) that similar convergence estimates as for FEM with piecewise linear basis functions hold. In particular, if the solution u is in $H^2(\Xi)$, one obtains linear convergence in the H^1 -norm and quadratic convergence in the L^2 -norm with respect to the mesh parameter h . However, due to the non-conformity of the test and trial spaces the FVM cannot be generalised to higher polynomial degrees of the ansatz spaces on unstructured grids in a straightforward manner as it is the case for FEM.

4.2 The Scharfetter-Gummel Discretisation

In problems of the form (4.1), where convection is dominant, stabilisation schemes are necessary to avoid spurious oscillations of the numerical solution. In the area of device simulation by means of the FVM, the most prominent technique is the *Scharfetter-Gummel scheme* [41]. It is an upwind-scheme, which means that it takes the direction of convective transport into account. In this section the Scharfetter-Gummel scheme is derived and we will see in the next section that the entropy considerations for the SHE of the BTE leads to a stabilisation scheme that can be seen as a generalisation of the Scharfetter-Gummel scheme.

We consider

$$\nabla \cdot (\nabla u + u \nabla \psi) = 0, \quad (4.16)$$

equipped with homogeneous Neumann boundary conditions for a given scalar quantity ψ . The standard FVM transforms this equation into

$$\sum_{B \in \mathcal{B}} \int_{\partial B} v (\nabla u + u \nabla \psi) \cdot \mathbf{n} \, ds = 0 \quad \forall v \in \mathcal{P}_0(\mathcal{B}). \quad (4.17)$$

The assembly of the system of discretised equation requires the value of the integrand at the interface ∂B . However, typically both ψ and u are discretised using a triangulation \mathcal{T} ,

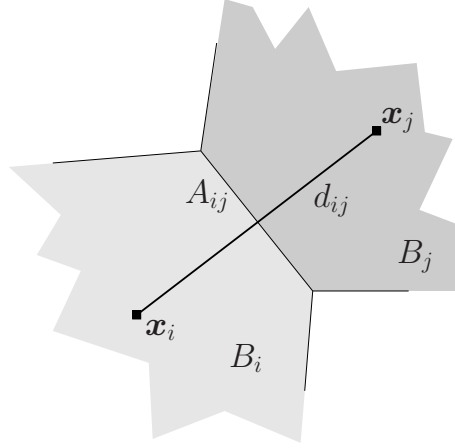


Figure 4.2: The length between vertices \mathbf{x}_i and \mathbf{x}_j is labelled d_{ij} , while the area of the interface between boxes B_i and B_j is A_{ij} .

hence the values at the interfaces have to be interpolated from values known at the vertices of the triangulation \mathcal{T} . To be more precise, let us introduce the following notation for a dual grid constructed using the barycenter method (cf. Fig. 4.2):

$$A_{ij} := \text{vol}(\partial B_i \cap \partial B_j) , \quad (4.18)$$

$$d_{ij} := |\mathbf{x}_i - \mathbf{x}_j| . \quad (4.19)$$

Hence, if we consider (4.16) along the edge connecting \mathbf{x}_i and \mathbf{x}_j with arc length parameter ξ along the edge, we have

$$\frac{du(\xi)}{d\xi} + u(\xi) \frac{d\psi(\xi)}{d\xi} = c_{ij} \quad (4.20)$$

with boundary conditions $u(0) = u(\mathbf{x}_i) =: u_i$ and $u(d_{ij}) = u(\mathbf{x}_j) =: u_j$. The constant c_{ij} arises from integration of the (now one-dimensional) divergence. Note that since c_{ij} provides an additional degree of freedom, it is not contradictory to specify two boundary conditions for the first-order ordinary differential equation (4.20). If we assume $\psi' := d\psi/d\xi$ to be constant along the edge with value $\psi'_{ij} \neq 0$, the solution of (4.20) is given as

$$u(\xi) = \frac{c_{ij}}{\psi'_{ij}} + A \exp(-\psi'_{ij} \xi) , \quad (4.21)$$

where A and c_{ij} have to be chosen such that the boundary conditions are fulfilled. Some calculations lead to the simple form

$$u(\xi) = (1 - g_{ij}(\xi))u_i + g_{ij}(\xi)u_j , \quad (4.22)$$

with growth function

$$g_{ij}(\xi) = \frac{1 - \exp(-\psi'_{ij} \xi)}{1 - \exp(-\psi'_{ij} d_{ij})} . \quad (4.23)$$

Since we assumed that the interface between B_i and B_j lies on the perpendicular bisector of the edge connecting \mathbf{x}_i and \mathbf{x}_j , we need to evaluate (4.22) and its derivative at $\xi = d_{ij}/2$ for an evaluation of (4.17):

$$u(d_{ij}/2) = \frac{u_i}{1 + \exp(\psi'_{ij}d_{ij}/2)} + \frac{u_j}{1 + \exp(-\psi'_{ij}d_{ij}/2)} , \quad (4.24)$$

$$\frac{du(d_{ij}/2)}{d\xi} = \frac{\psi'_{ij}}{2} \frac{u_j - u_i}{\sinh(\psi'_{ij}d_{ij}/2)} . \quad (4.25)$$

Thus, the contribution of the interface $\partial B_i \cap \partial B_j$ to (4.17) using the midpoint rule for integration can be written as

$$\begin{aligned} \int_{\partial B_i \cap \partial B_j} v(\nabla u + u\nabla\psi) \cdot \mathbf{n} \, ds &\approx A_{ij} v\left(\frac{\mathbf{x}_i + \mathbf{x}_j}{2}\right) \left[\nabla u\left(\frac{\mathbf{x}_i + \mathbf{x}_j}{2}\right) + u\left(\frac{\mathbf{x}_i + \mathbf{x}_j}{2}\right) \nabla\psi \right] \\ &= A_{ij} \left[\frac{\psi'_{ij}}{2} \frac{u_j - u_i}{\sinh(\psi'_{ij}d_{ij}/2)} \right. \\ &\quad \left. + \left(\frac{u_i}{1 + \exp(\psi'_{ij}d_{ij}/2)} + \frac{u_j}{1 + \exp(-\psi'_{ij}d_{ij}/2)} \right) \psi'_{ij} \right] \\ &= \frac{A_{ij}}{d_{ij}} \left[u_j \text{Bern}(-\psi'_{ij}d_{ij}) - u_i \text{Bern}(\psi'_{ij}d_{ij}) \right] , \end{aligned} \quad (4.26)$$

where

$$\text{Bern}(x) = \frac{x}{\exp(x) - 1} \quad (4.27)$$

is the *Bernoulli function* depicted in Fig. 4.3. The term ψ'_{ij} , which has been assumed constant along the edge, is approximated by a simple discrete derivative of the form

$$\psi'_{ij} \approx \frac{\psi(\mathbf{x}_j) - \psi(\mathbf{x}_i)}{d_{ij}} . \quad (4.28)$$

The appealing property of the Scharfetter-Gummel scheme is the following: Depending on the sign and magnitude of the term $\psi'_{ij}d_{ij}$, either u_j or u_i is primarily used for the approximation of the convective flux through the interface between the two boxes B_i and B_j in (4.26). In case that $\psi'_{ij}d_{ij} \approx 0$, equal weight is assigned to u_i and u_j , thus we obtain the usual central difference scheme for (4.16) with negligible convection term. Thus, the Scharfetter-Gummel scheme is an upwind scheme that takes the hyperbolic nature of the convective term in (4.16) into account.

4.3 Discretisation of the SHE of the BTE

In this section a discretisation scheme for the SHE equation (3.30) is derived. A natural requirement on a discretisation scheme is that it reflects mathematical and physical properties included in the continuous model. For the purpose of predictive device simulation, local charge conservation is such a major requirement. Additionally, entropy considerations in Section 3.3 have lead to the requirement that entropy is preserved by the free streaming

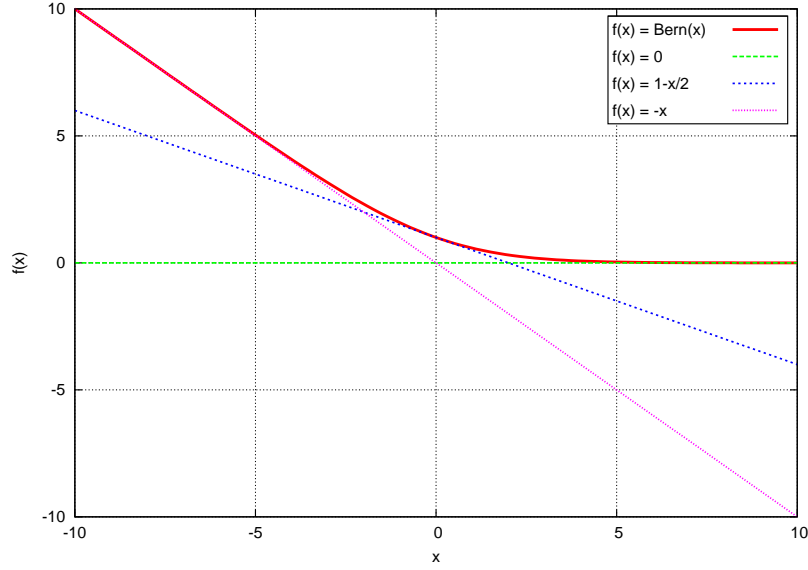


Figure 4.3: The Bernoulli function $\text{Bern}(x)$.

operator. Unfortunately, these two requirements are somewhat contradictory: Building the zero order moment of the steady-state BTE with respect to the wave vector yields

$$\int_{\mathbb{R}^3} \frac{\partial f}{\partial t} + \nabla_{\mathbf{x}} \cdot (\mathbf{v}(\mathbf{k})f) + \frac{1}{\hbar} \nabla_{\mathbf{k}} \cdot (\mathbf{F}f) d\mathbf{k} = \int_{\mathbb{R}^3} Q\{f\} d\mathbf{k} . \quad (4.29)$$

We assume that the collision operator is charge-conserving, thus the integral on the right hand side vanishes. The third term on the left hand side can be cast into an integral over the boundary of the \mathbf{k} -space due to Gauss' theorem and thus vanishes since the distribution function f decays to zero as $|\mathbf{k}|$ goes to infinity. Therefore we arrive at the *continuity equation*

$$\underbrace{\frac{\partial}{\partial t} \int_{\mathbb{R}^3} f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k}}_{\sim n(\mathbf{x}, t)} + \nabla_{\mathbf{x}} \cdot \underbrace{\int_{\mathbb{R}^3} \mathbf{v}(\mathbf{k}) f(\mathbf{x}, \mathbf{k}, t) d\mathbf{k}}_{\sim \mathbf{j}(\mathbf{x}, t)} = 0 . \quad (4.30)$$

This equation should hold locally such that we are able to apply a discrete version of Gauss' theorem locally over every sub-mesh. However, if we apply a finite element discretisation in spatial direction, entropy considerations suggest a discretisation using a weighted scalar product, which would automatically not be locally conservative. Similar problems arise with difference schemes.

As a compromise between entropy preservation and charge conservation, the following strategy was suggested by Ringhofer [37]: The BTE is formally split into even and odd parts as in Section 3.2. Discretisation schemes for the even part are designed in such a way that even momenta are locally conserved. The scheme for the odd part is then chosen such that the whole scheme possesses entropy properties as derived in Section 3.3. Such a compromise is acceptable, since odd order moments are not conserved anyway. Consequently, they have to be viewed rather as constitutive relations [37].

We now turn to the SHE equations (3.30) and show that charge is conserved: Integration of (3.30) over energy leads to

$$\int_0^\infty \frac{\partial f_{l,m}}{\partial t} + \nabla_{\mathbf{x}} \cdot \mathbf{j}_{l,m} + \mathbf{F} \cdot \left(\frac{\partial \mathbf{j}_{l,m}}{\partial \varepsilon} - \mathbf{\Gamma}_{l,m} \right) d\varepsilon = 0 , \quad (4.31)$$

where we have used that the collision operator is charge conserving. Since according to (3.4) the current density is obtained from $\mathbf{j}_{0,0}$, it is sufficient to consider the case $l = m = 0$ only. Since $Y_{0,0}$ is constant, $\mathbf{\Gamma}_{0,0,\nu',m'}$ (cf. (3.48)) vanishes. Moreover, the integral over $\mathbf{F} \cdot \partial \mathbf{j}_{0,0} / \partial \varepsilon$ is zero, because $\mathbf{j}_{0,0}$ vanishes at zero and infinite energy [33]. Thus, with the charge density (3.4) and the particle current density (3.11) we obtain

$$\frac{\partial n}{\partial t} + \nabla_{\mathbf{x}} \cdot \mathbf{j} = 0 \quad (4.32)$$

and therefore the discrete system of equations will conserve the particle density exactly if a discretisation is carried out using FVM for even values of l .

For the discretisation of the SHE equations for odd l , we multiply the BTE with

$$H(\mathbf{x}, \varepsilon) = \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right) , \quad (4.33)$$

to obtain

$$H(\mathbf{x}, \varepsilon) \frac{\partial f}{\partial t} + H(\mathbf{x}, \varepsilon) \mathbf{v}(\mathbf{k}) \nabla_{\mathbf{x}} f + H(\mathbf{x}, \varepsilon) \frac{1}{\hbar} \mathbf{F} \nabla_{\mathbf{k}} f = H(\mathbf{x}, \varepsilon) Q\{f\} . \quad (4.34)$$

For notational convenience, we write H instead of $H(\mathbf{x}, \varepsilon)$ in the following. Since

$$\mathbf{v}(\mathbf{k}) \nabla_{\mathbf{x}} (Hf) + \frac{1}{\hbar} \mathbf{F} \nabla_{\mathbf{k}} (Hf) = H \mathbf{v}(\mathbf{k}) \nabla_{\mathbf{x}} f + H \frac{1}{\hbar} \mathbf{F} \nabla_{\mathbf{k}} f , \quad (4.35)$$

and H depends on \mathbf{k} only over the energy ε , one obtains the SHE equation

$$H \frac{\partial f_{l,m}}{\partial t} + \nabla_{\mathbf{x}} \cdot (H \mathbf{j}_{l,m}) + \mathbf{F} \cdot \left(\frac{\partial H \mathbf{j}_{l,m}}{\partial \varepsilon} - H \mathbf{\Gamma}_{l,m} \right) = H Q_{l,m}\{f\} , \quad (4.36)$$

which can now be discretised using finite volumes or finite differences.

To unify the treatment of even and odd parts of the distribution function, we define

$$H_l(\mathbf{x}, \varepsilon) = \begin{cases} 1 & \text{for even } l , \\ \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right) & \text{for odd } l , \end{cases} \quad (4.37)$$

so that we arrive at the system of equations

$$H_l \frac{\partial f_{l,m}}{\partial t} + \nabla_{\mathbf{x}} \cdot (H_l \mathbf{j}_{l,m}) + \mathbf{F} \cdot \left(H_l \frac{\partial \mathbf{j}_{l,m}}{\partial \varepsilon} - H_l \mathbf{\Gamma}_{l,m} \right) = H_l Q_{l,m}\{f\} , \quad (4.38)$$

for $l = 0, 1, \dots$ and $m = -l, \dots, l$, or in divergence form

$$H_l \frac{\partial f_{l,m}}{\partial t} + \tilde{\nabla} \cdot (H_l \tilde{\mathbf{j}}_{l,m}) - H_l \mathbf{F} \cdot \mathbf{\Gamma}_{l,m} = H_l Q_{l,m}\{f\} . \quad (4.39)$$

A finite volume scheme can now be obtained as outlined in Section 4.1: Multiplication with a test function $v \in \mathcal{P}_0(\mathcal{B})$ (which is not be confused with the velocity $\mathbf{v}(\mathbf{k})$), integration over the problem domain G and application of Gauss' theorem locally for each box B_i yields

$$\int_{B_i} H_l \frac{\partial f_{l,m}}{\partial t} v \, d\boldsymbol{\xi} + \int_{\partial B_i} H_l \tilde{\mathbf{j}}_{l,m} v \, d\mathbf{A} - \int_{B_i} \mathbf{F} H_l \boldsymbol{\Gamma}_{l,m} v \, d\boldsymbol{\xi} = \int_{B_i} H_l Q_{l,m} \{\tilde{f}\} v \, d\boldsymbol{\xi} , \quad (4.40)$$

for all $v \in \mathcal{P}_0(\mathcal{B})$, where $\boldsymbol{\xi} = (\mathbf{x}, \varepsilon)$ is the position vector in the four-dimensional spatial/energy space. The expansion of the generalised current continuity $\tilde{\mathbf{j}}_{l,m}$ in (3.46) and the expansion of $\boldsymbol{\Gamma}_{l,m}$ in (3.48) can now be inserted into (4.40) to obtain a system of equations in the unknown variables $f_{l,m}$. For a triangulation \mathcal{T} of the problem domain, a discretisation of $f_{l,m}$ is carried out in the form

$$f_{l,m} = \sum_{i=1}^N f_{l,m;i}(t) \varphi_i(\boldsymbol{\xi}) , \quad (4.41)$$

where N denotes the number of vertices of the triangulation and the basis φ_i was introduced in (4.11). The coefficients $f_{l,m;i}(t)$ are the unknowns of the resulting system of equations without discretisation of the time variable. Time discretisation is not further considered in this thesis, however a straight-forward backward Euler scheme can be derived in the usual way. Substitution of the expressions for $\tilde{\mathbf{j}}_{l,m}$ and $\boldsymbol{\Gamma}_{l,m}$, use of the velocity randomising scattering operator (3.27) and a substitution of (4.41) on the left hand side of (4.40) yields for a box B_i

$$\begin{aligned} & \int_{B_i} H_l \sum_{i'=1}^N \frac{\partial f_{l,m;i'}}{\partial t} \varphi_{i'}(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\ & + \int_{\partial B_i} H_l \sum_{i'=1}^N \sum_{l'=0}^L \sum_{m'=-l'}^{l'} \begin{pmatrix} \mathbf{v}_{l,m,l',m'} \\ \mathbf{F} \cdot \mathbf{v}_{l,m,l',m'} \end{pmatrix} f_{l',m';i'} \varphi_{i'}(\boldsymbol{\xi}) \, d\mathbf{A} \\ & - \int_{B_i} \mathbf{F} H_l \sum_{i'=1}^N \sum_{l'=0}^L \sum_{m'=-l'}^{l'} \boldsymbol{\Gamma}_{l,m,l',m'} f_{l',m';i'} \varphi_{i'}(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\ & = \frac{1}{Y_{0,0}} \sum_{\eta} \int_{B_i} H_l \left[g_{l,m}(\varepsilon) \sigma_{\eta}(\mathbf{x}, \varepsilon \pm \hbar \omega_{\eta}, \varepsilon) \sum_{i'=1}^N f_{0,0;i'} \varphi_{i'}(\boldsymbol{\xi} \pm \hbar \omega_{\eta}) \right. \\ & \quad \left. - \sum_{i'=1}^N f_{l,m;i'} \varphi_{i'}(\boldsymbol{\xi}) \sigma_{\eta}(\mathbf{x}, \varepsilon, \varepsilon \mp \hbar \omega_{\eta}) g_{0,0}(\varepsilon \mp \hbar \omega_{\eta}) \right] d\boldsymbol{\xi} , \end{aligned} \quad (4.42)$$

where

$$\boldsymbol{\xi} = \begin{pmatrix} \mathbf{x} \\ \varepsilon \end{pmatrix} , \quad \boldsymbol{\omega}_{\eta} = \begin{pmatrix} \mathbf{0} \\ \omega_{\eta} \end{pmatrix}$$

and we have dropped the test function v , since by the use of linear independence of the basis X in (4.10) it suffices to consider the single basis function which evaluates to one in the box B_i . The coefficients

$$\mathbf{v}_{l,m,l',m'} = \mathbf{v}_{l,m,l',m'}(\varepsilon) = \int_{\Omega} Y_{l,m} \mathbf{v}(\mathbf{k}(\varepsilon, \theta, \varphi)) Y_{l',m'} \, d\Omega , \quad (4.43)$$

$$\boldsymbol{\Gamma}_{l,m,l',m'} = \boldsymbol{\Gamma}_{l,m,l',m'}(\varepsilon) = \int_{\Omega} \frac{1}{\hbar |\mathbf{k}(\varepsilon, \theta, \varphi)|} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_{\varphi} \right) Y_{l',m'} \, d\Omega \quad (4.44)$$

depend on energy, thus either numerical quadrature or a projection onto lower order polynomials is necessary for the assembly of the system (4.42). This energy dependence stems from the band structure under consideration. In the case of spherical bands such as the multi-band structure presented in Section 2.3, the energy depends on the modulus of the wave vector only, thus we can write

$$\mathbf{v}_{l,m,l',m'}(\varepsilon) = \tilde{v}(\varepsilon) \underbrace{\int_{\Omega} Y_{l,m} \mathbf{e}_{\varepsilon} Y_{l',m'} d\Omega}_{:=\mathbf{a}_{l,m,l',m'}}, \quad (4.45)$$

$$\mathbf{\Gamma}_{l,m,l',m'}(\varepsilon) = \frac{1}{\hbar \tilde{k}(\varepsilon)} \underbrace{\int_{\Omega} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_{\varphi} \right) Y_{l',m'} d\Omega}_{:=\mathbf{b}_{l,m,l',m'}} \quad (4.46)$$

and the energy dependence is fully decoupled from the dependence on the spherical harmonics. Moreover, the coefficients $\mathbf{a}_{l,m,l',m'}$ and $\mathbf{b}_{l,m,l',m'}$ can be evaluated algebraically, which is done in the next chapter. For the inclusion of full-band effects, a numerical integration is necessary to obtain $\mathbf{v}_{l,m,l',m'}$ and $\mathbf{\Gamma}_{l,m,l',m'}$. Since these quantities do not depend on the selected triangulation of the problem domain of (4.42), they can again be evaluated in a preprocessing step, leading to no additional effort during the assembly of the system. Jungemann et. al. [24] suggested a projection of the dependence on the wave vector onto spherical harmonics; as we will see later on, such a projection is an attractive way to approximate the integrals in (4.43) and (4.44).

4.4 Rectangular Grids in Two Dimensions

Coming back to (4.42), we notice that the integrands of the discretised system carry an exponential in case that the major index l is odd. This complicates the numerical integration of the integrals, especially in the presence of a large electric field. It is therefore appealing to consider structured grids, especially rectangular grids, where the dual grid is again a rectangular grid, and derive analytical expressions for the integral terms. A similar approach was used by Jungemann et. al. [24], which additionally incorporates some simplifications of the integrands in (4.42). In the following we derive such analytical expressions for the Galerkin method that has lead to the system (4.42).

In one spatial dimension and one energy dimension, a Cartesian product grid (x_i, ε_j) with N_x and N_{ε} grid points in x - and ε -direction respectively can be used. The dual grid of this rectangular grid is again a rectangular grid with nodes located at

$$\begin{pmatrix} x_{i+1/2} \\ \varepsilon_{j+1/2} \end{pmatrix} := \begin{pmatrix} \frac{x_i + x_{i+1}}{2} \\ \frac{\varepsilon_j + \varepsilon_{j+1}}{2} \end{pmatrix}, \quad i = 1, \dots, N_x - 1, \quad j = 1, \dots, N_{\varepsilon} - 1. \quad (4.47)$$

The electrostatic potential ψ is linearly interpolated between the grid nodes:

$$\psi_{i+1/2} = \frac{\psi_i + \psi_{i+1}}{2}, \quad i = 1, \dots, N_x - 1. \quad (4.48)$$

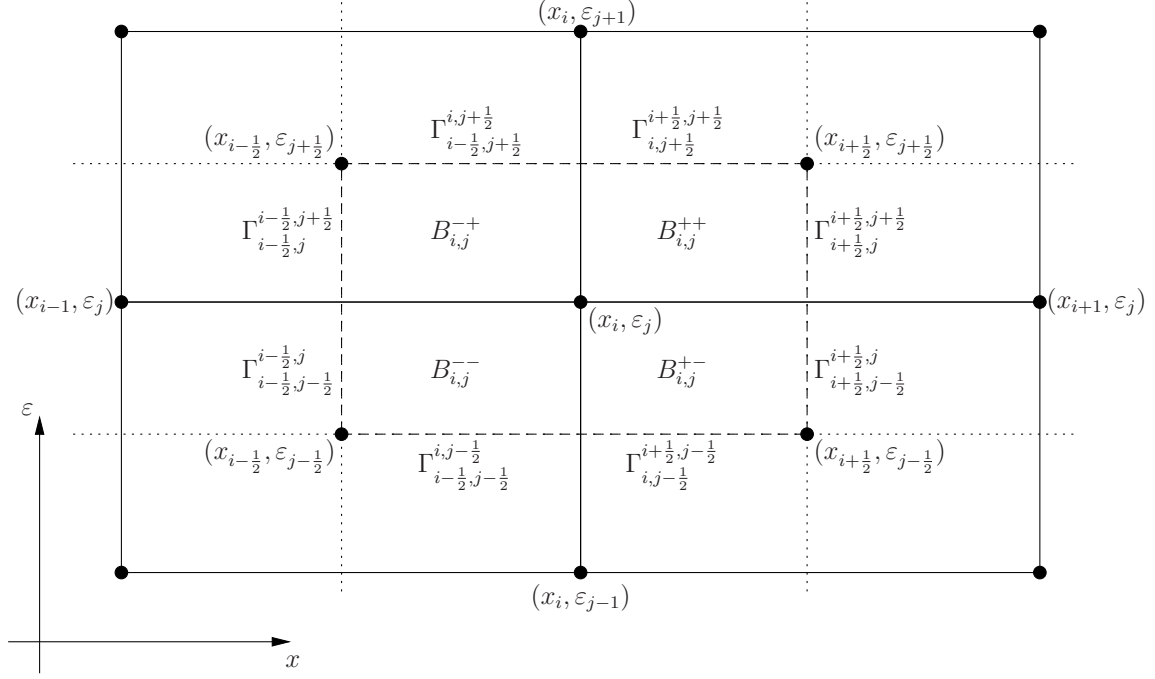


Figure 4.4: Labels associated with a rectangular box $B_{i,j}$ centered at (x_i, ε_j) .

A nodal basis on the reference rectangle with opposite corners at $(0,0)$ and $(1,1)$ is

$$\varphi_1(x, \varepsilon) = (1-x)(1-\varepsilon), \quad (4.49)$$

$$\varphi_2(x, \varepsilon) = x(1-\varepsilon), \quad (4.50)$$

$$\varphi_3(x, \varepsilon) = (1-x)\varepsilon, \quad (4.51)$$

$$\varphi_4(x, \varepsilon) = x\varepsilon. \quad (4.52)$$

By affine transformations these functions can be mapped to the rectangle of interest. Even though these nodal basis functions are not linear functions per se, they are linear along parameter lines. On structured grids, this means that each basis function is linear along interfaces between rectangles.

To investigate the discretisation on rectangular grids further, let us introduce the following notations (cf. Fig. 4.4):

- $B_{i,j}$: Box centered at (x_i, ε_j)
- $B_{i,j}^{++}, B_{i,j}^{+-}, B_{i,j}^{-+}, B_{i,j}^{--}$: Decomposition of $B_{i,j}$ into boxes spanned by (x_i, ε_j) and points $(x_{i\pm 1/2}, \varepsilon_{j\pm 1/2})$ respectively.
- $\Gamma_{i_1, j_1}^{i_2, j_2}$: Arc between $(x_{i_1}, \varepsilon_{j_1})$ and $(x_{i_2}, \varepsilon_{j_2})$.
- $\varphi_{i,j}$: Nodal basis function associated with (x_i, ε_j) .
- $f_{l,m,i,j}$: Spherical harmonics expansion term of $Y_{l,m}$ and $\varphi_{i,j}$.

Before we derive analytical formulas for the integrals in (4.42) one after another, we first consider the integral

$$\int_a^b \exp(\gamma x) \, dx \quad (4.53)$$

for arbitrary values of a , b and $\gamma \neq 0$. With integration per parts one finds

$$\int_a^b \exp(\gamma x) \, dx = \frac{1}{\gamma} [\exp(\gamma b) - \exp(\gamma a)] . \quad (4.54)$$

While the integral on the left hand side also allows for an evaluation in case $\gamma = 0$, the right hand side does not. A numerical evaluation of the right hand side is also dangerous for $|\gamma| \ll 1$, because numerical noise enters due to the subtraction of the two exponential terms of similar magnitude, which is then amplified by a factor of $1/\gamma$. By use of the Bernoulli function (4.27) we rewrite

$$\begin{aligned} \frac{1}{\gamma} [\exp(\gamma b) - \exp(\gamma a)] &= (b - a) \exp(a) \frac{\exp(\gamma(b - a)) - 1}{\gamma(b - a)} \\ &= (b - a) \exp(a) \text{Bern}(\gamma(b - a)) \end{aligned} \quad (4.55)$$

and have now obtained a numerically stable expression, because the continuous extension of the Bernoulli function $\text{Bern}(\cdot)$ at $x = 0$ can be included in the program code. Similar numerical stabilisations have been applied to the formulae given in the following. The full calculations are often tedious but follow a common pattern similar to (4.55) and are thus omitted.

- $\int_{B_{i,j}} H_l \sum_{i'=1}^{N_x} \sum_{j'=1}^{N_\varepsilon} \frac{\partial f_{l,m;i',j'}}{\partial t} \varphi_{i',j'}(\xi) \, d\xi$: In case that l is even, the integrand is piecewise quadratic with respect to the integration variables and the integral can be evaluated easily for each box. For $B_{i,j}^{++}$ and a factorisation of the basis function of the form $\varphi(x, \varepsilon) = \varphi_x(x) \varphi_\varepsilon(\varepsilon)$ (c.f. (4.49)-(4.52)) we find

$$\begin{aligned} \int_{B_{i,j}^{++}} \varphi \, d\xi &= \int_{x_i}^{x_{i+1/2}} \varphi_x(x) \, dx \times \int_{\varepsilon_i}^{\varepsilon_{i+1/2}} \varphi_\varepsilon(\varepsilon) \, d\varepsilon \\ &= \text{vol}(B_{i,j}^{++}) [\varphi_x(x_i) + \varphi_x(x_{i+1/2})] [\varphi_\varepsilon(\varepsilon_j) + \varphi_\varepsilon(\varepsilon_{j+1/2})] / 4 , \end{aligned} \quad (4.56)$$

where $\text{vol}(B_{i,j}^{++}) = (x_{i+1/2} - x_i)(\varepsilon_{j+1/2} - \varepsilon_j)$. For the boxes $B_{i,j}^{+-}$, $B_{i,j}^{-+}$ and $B_{i,j}^{--}$ similar expressions are obtained.

For odd l the integrand is made up from a polynomial times an exponential due to the stabilisation term H_l . A standard technique is now to rely on numerical quadrature rules, but one can also evaluate the resulting integrals analytically: Integration over the box $B_{i,j}^{++}$ with the shorthand notations $x^+ := x_{i+1/2}$, $x^- := x_i$, $\varepsilon^+ := \varepsilon_{i+1/2}$ and

$\varepsilon^- := \varepsilon_i$ yields

$$\begin{aligned}
\int_{B_{i,j}^{++}} H_l \varphi \, d\boldsymbol{\xi} &= \int_{x_i}^{x_{i+\frac{1}{2}}} \int_{\varepsilon_j}^{\varepsilon_{j+\frac{1}{2}}} \exp\left(\frac{\varepsilon + q\psi(x)}{k_B T}\right) \varphi(x, \varepsilon) \, d\boldsymbol{\xi} \\
&= \int_{x_i}^{x_{i+\frac{1}{2}}} \exp\left(\frac{q\psi(x)}{k_B T}\right) \varphi_x(x) \, dx \times \int_{\varepsilon_j}^{\varepsilon_{j+\frac{1}{2}}} \exp\left(\frac{\varepsilon}{k_B T}\right) \varphi_\varepsilon(\varepsilon) \, d\varepsilon \\
&= \exp\left(\frac{\psi(x_i)}{V_T}\right) (x^+ - x^-) \left[\frac{\varphi_x(x^+) + (\varphi_x(x^+) - \varphi_x(x^-))\mathcal{C}(\Delta_\psi)}{\text{Bern}(\Delta_\psi)} \right] \\
&\quad \times \exp\left(\frac{\varepsilon_j}{k_B T}\right) (\varepsilon^+ - \varepsilon^-) \left[\frac{\varphi_\varepsilon(\varepsilon^+) + (\varphi_\varepsilon(\varepsilon^+) - \varphi_\varepsilon(\varepsilon^-))\mathcal{C}(\Delta_\varepsilon)}{\text{Bern}(\Delta_\varepsilon)} \right]
\end{aligned} \tag{4.57}$$

where

$$\mathcal{C}(x) = \frac{\text{Bern}(x) - 1}{x} = \frac{x - \exp(x) + 1}{x(\exp(x) - 1)} \tag{4.58}$$

and the labels

$$V_T = \frac{k_B T}{q}, \quad \Delta_\psi = \frac{\psi_{i+1/2} - \psi_i}{V_T}, \quad \Delta_\varepsilon = \frac{\varepsilon_{j+1/2} - \varepsilon_j}{k_B T}$$

were used. Let us have a closer look at the stabilisation function $\mathcal{C}(x)$:

$$\begin{aligned}
\mathcal{C}(x) &= \frac{x - \exp(x) + 1}{x(\exp(x) - 1)} \\
&= -\frac{\sum_{k=2}^{\infty} x^k / k!}{\sum_{k=2}^{\infty} x^k / (k-1)!} .
\end{aligned} \tag{4.59}$$

Thus, we deduce the following asymptotic behaviour:

$$\mathcal{C}(x) \sim \begin{cases} -1/x, & x \gg 1 \\ -1/2 + x/12, & |x| \ll 1 \\ -1, & x \ll -1 \end{cases} . \tag{4.60}$$

If we use these asymptotics for the numerator of the x -dependent part of (4.57), we obtain

$$\varphi_x(x^+) + (\varphi_x(x_{x^+}) - \varphi_x(x^-))\mathcal{C}(\Delta_\psi) \sim \begin{cases} \varphi_x(x_{x^+}), & \Delta_\psi \gg 1 \\ (\varphi_x(x^+) + \varphi_x(x^-))/2, & |\Delta_\psi| \ll 1 \\ \varphi_x(x^-), & \Delta_\psi \ll -1 \end{cases} ,$$

which can be interpreted as a selection of an appropriate integration weight for φ_x depending on the variation of ψ .

- $\int_{\partial B_{i,j}} H_l \left(\begin{smallmatrix} \mathbf{F} \cdot \mathbf{v}_{l,m,l',m'} \\ \mathbf{v}_{l,m,l',m'} \end{smallmatrix} \right) f_{l',m';i',j'} \varphi_{i',j'}(\boldsymbol{\xi}) \, d\mathbf{A}$ (summed over i', j', l' and m'): The first observation is that test functions are linear along each edge of the rectangular grid. Moreover, since the rectangular grid has edges aligned to the coordinate directions, the ε - and x -component of the result can be obtained by integration along x

and ε respectively. Let us consider the ε component on one boundary segment, say $\Gamma_{i,j+1/2}^{i+1/2,j+1/2}$, first, where we keep only those terms that depend on the integration variable. In case that l is even, we have

$$\int_{x_i}^{x_{i+\frac{1}{2}}} H_l \mathbf{F} \varphi \, dx = -q \nabla \psi \int_{x_i}^{x_{i+\frac{1}{2}}} \varphi \, dx = -q(\psi_{i+\frac{1}{2}} - \psi_i)(\varphi_{i+\frac{1}{2}} - \varphi_i)/2, \quad (4.61)$$

with $\varphi_{i+\frac{1}{2}}$ and φ_i being the values of the test function evaluated at $x_{i+\frac{1}{2}}$ and x_i and similarly for the electrostatic potential ψ . Note that, since the electrostatic potential is interpolated linearly, the gradient is constant along the integration interval.

For odd l a calculation similar to (4.57) using the shorthand notation $\psi'_{i',i''} := (\psi_{i'} - \psi_{i''})/V_T$ yields

$$\begin{aligned} \int_{x_i}^{x_{i+\frac{1}{2}}} H_l \mathbf{F} \varphi \, dx &= -q \nabla \psi \exp\left(\frac{\varepsilon}{k_B T}\right) \int_{x_i}^{x_{i+\frac{1}{2}}} \exp\left(\frac{\psi(x)}{V_T}\right) \varphi \, dx \\ &= -q \psi'_{i+\frac{1}{2},i} \exp\left(\frac{\varepsilon + q\psi_i}{k_B T}\right) \left[\frac{\varphi_{i+\frac{1}{2}} + (\varphi_{i+\frac{1}{2}} - \varphi_i) \mathcal{C}(\psi'_{i+\frac{1}{2},i})}{\text{Bern}(\psi'_{i+\frac{1}{2},i})} \right], \end{aligned} \quad (4.62)$$

where we have used the notation $\varphi_{i'} := \varphi(x_{i'})$ as well as the Bernoulli function and (4.58) to avoid numerical instabilities. The contribution of other edges can be obtained by replacing the indices i and $i + 1/2$ appropriately.

Integration along edges in energy direction is carried out similarly: We interpolate the energy-dependence of the velocity \mathbf{v} piecewise linearly¹. For even l the integrand then is a product of two linear functions and can be conveniently evaluated using Simpson's rule:

$$\int_{\varepsilon_j}^{\varepsilon_{j+\frac{1}{2}}} \mathbf{v} \varphi \, d\varepsilon = \frac{\varepsilon_{j+\frac{1}{2}} - \varepsilon_j}{6} \left(v_j \varphi_j + (v_j + v_{j+\frac{1}{2}})(\varphi_j + \varphi_{j+\frac{1}{2}}) + v_{j+\frac{1}{2}} \varphi_{j+\frac{1}{2}} \right), \quad (4.63)$$

where we have set $v_{j'} := v(\varepsilon_{j'})$.

In case that l is odd, the integrand consists of a product of an exponential and two linear functions. Another lengthy calculation gives

$$\begin{aligned} \int_{\varepsilon_j}^{\varepsilon_{j+\frac{1}{2}}} \exp\left(\frac{\varepsilon + q\psi}{k_B T}\right) \mathbf{v} \varphi \, d\varepsilon &= \exp\left(\frac{\varepsilon_j + q\psi_i}{k_B T}\right) \frac{k_B T}{\text{Bern}(\varepsilon'_{j+\frac{1}{2},j})} \left\{ \right. \\ &\quad \left[\exp\left(\varepsilon'_{j+\frac{1}{2},j}\right) v_{j+\frac{1}{2}} \varphi_{j+\frac{1}{2}} - v_j \varphi_j \right] \text{Bern}(\varepsilon'_{j+\frac{1}{2},j}) \\ &\quad - 2(v_{j+\frac{1}{2}} - v_j)(\varphi_{j+\frac{1}{2}} - \varphi_j) \mathcal{C}(\varepsilon'_{j+\frac{1}{2},j}) \\ &\quad \left. - (v_{j+\frac{1}{2}} - v_j) \varphi_{j+\frac{1}{2}} - (\varphi_{j+\frac{1}{2}} - \varphi_j) v_{j+\frac{1}{2}} \right\}, \end{aligned} \quad (4.64)$$

with $\varepsilon'_{j',j''} := (\varepsilon_{j'} - \varepsilon_{j''})/(k_B T)$. Similar expressions for the remaining edge segments are again obtained by replacing the indices appropriately.

¹Note that in case of parabolic bands we have $v \sim \sqrt{\varepsilon}$, thus a linear interpolation can only be justified away from $\varepsilon = 0$.

The integral over the full boundary of the box can now be assembled by summing the individual contributions appropriately. Some care has to be taken because some edges have a surface normal pointing into negative x - or ε -direction, thus these contributions have to be taken with a minus sign.

- $\int_{B_{i,j}} H_l \mathbf{F} \cdot \mathbf{\Gamma}_{l,m,l',m'} f_{l',m';i',j'} \varphi_{i',j'} d\xi$ (summed over i', j', l' and m'): We decompose the box into four sub-boxes $B_{i,j}^{++}$, $B_{i,j}^{+-}$, $B_{i,j}^{-+}$ and $B_{i,j}^{--}$. Since \mathbf{F} is a function of the spatial variable x , $\mathbf{\Gamma}$ is a function of the energy variable and φ can be written as a product of a function that depends on x and a function that depends on ε (cf. (4.49)-(4.52)), the double-integral over the box can be rewritten as a product of two single integrals. More explicitly, writing $\varphi(x, \varepsilon) = \varphi_x(x) \varphi_\varepsilon(\varepsilon)$, then

$$\int_{B_{i,j}^{++}} \mathbf{F}(\mathbf{x}) \cdot \mathbf{\Gamma}(\varepsilon) \varphi d\xi = - \int_{x_i}^{x_{i+1/2}} q \nabla \psi \varphi_x dx \times \int_{\varepsilon_i}^{\varepsilon_{i+1/2}} \Gamma \varphi_\varepsilon d\varepsilon \quad (4.65)$$

for even l . We interpolate Γ linearly², hence the first integral on the right hand side was evaluated in (4.61) and the second in (4.63) where one has to replace \mathbf{v} with $\mathbf{\Gamma}$.

In case that l is odd,

$$\begin{aligned} \int_{B_{i,j}^{++}} \exp\left(\frac{\varepsilon + q\psi}{k_B T}\right) \mathbf{F} \cdot \mathbf{\Gamma} \varphi d\xi = & - \int_{x_i}^{x_{i+1/2}} q \nabla \psi \exp\left(\frac{q\psi}{k_B T}\right) \varphi_x dx \\ & \times \int_{\varepsilon_i}^{\varepsilon_{i+1/2}} \exp\left(\frac{\varepsilon}{k_B T}\right) \Gamma \varphi_\varepsilon d\varepsilon . \end{aligned} \quad (4.66)$$

Again, the first term on the right hand side was already given by (4.62) and the second by (4.64) after relabelling \mathbf{v} with $\mathbf{\Gamma}$. Expressions for integration over $B_{i,j}^{+-}$, $B_{i,j}^{-+}$ and $B_{i,j}^{--}$ are obtained in the same way and finally the integral over the full box is obtained by summation.

- $\int_{B_i} H_l Q_{l,m}^{\text{VR}} \{f\} d\xi$: First we assume that the energies $\hbar\omega_\eta$ are multiples of the grid spacing $\Delta\varepsilon$ so that final scattering states are located at grid nodes. We additionally assume constant collision cross sections σ_η , hence we have to evaluate integrals of the form

$$\int_{B_{i,j}} H_l g(\varepsilon) \varphi(\xi) d\xi \quad \text{and} \quad \int_{B_{i,j}} H_l g(\varepsilon \pm \hbar\omega) \varphi(\xi) d\xi .$$

Again, we approximate the generalised density of states g piecewise linearly³, so that the shift by $\hbar\omega$ does not need to be considered separately. With a factorisation of the basis function φ the integral over the box $B_{i,j}^{++}$ becomes

$$\int_{B_{i,j}^{++}} H_l g(\varepsilon) \varphi(\xi) d\xi = \int_{x_i}^{x_{i+1/2}} H_l \varphi_x(x) dx \times \int_{\varepsilon_j}^{\varepsilon_{j+1/2}} H_l g(\varepsilon) \varphi_\varepsilon(\varepsilon) d\varepsilon . \quad (4.67)$$

We can use again (4.61) and (4.63) in case of even l and (4.62) and (4.64) for odd l , where appropriate relabelling of \mathbf{v} with g is required.

²In case of parabolic bands, $\mathbf{\Gamma} \sim 1/\sqrt{\varepsilon}$, thus a linear interpolation can only be justified away from $\varepsilon = 0$

³In case of parabolic bands, $g \sim \sqrt{\varepsilon}$ and the piecewise linear approximation is poor near zero energy

The piecewise linear interpolation of the quantities \mathbf{v} , $\mathbf{\Gamma}$ and g is not a good approximation near zero energy, because the derivatives of all three quantities is singular there. Moreover, $\mathbf{\Gamma}$ is unbounded near zero energy, but integrable. Thus, a hybrid strategy should be chosen: Near zero energy a computationally demanding numerical quadrature is carried out, while away from the singularities the piecewise linear interpolation is quite accurate and thus analytic formulae can be used to speed up the assembly process.

4.5 Boundary Conditions

Now that we are able to compute all integral terms by explicit formulae, the system of linear equations with unknowns coefficients $f_{l,m;i}$ (or $f_{l,m;i,j}$ in the case of rectangular grids in two dimensions) can be set up. The system matrix consists of all terms evaluated in the previous section, while the right hand side vector appears - at least at first sight - to be zero, since there are no terms in (4.42) that do not depend on the unknowns. However, nonzero entries on the right hand side vector show up due to boundary conditions:

- At cells near the boundary of the energy domain, out-scattering coefficients with final energies and in-scattering coefficients with initial energies outside the discretised energy interval are set to zero (cf. for example (3.25) and (3.26)). At the energy boundary $\varepsilon = \varepsilon_{\max}$, homogeneous Neumann boundary conditions are applied in the sense that in parts of a cell energies larger than ε_{\max} are ignored.

At zero energy, there are in fact two possible boundary conditions: If the generalised energy distribution function \tilde{f} is discretised, then all coefficients $f_{l,m}$ are zero at zero energy provided that the density of states is zero at zero energy. Even if the distribution function is discretised directly, the coefficients of the odd harmonics are zero at zero energy, provided that the wave vector vanishes there. This can be reasoned as follows: Consider the SHE of the distribution function at zero energy. Since \mathbf{k} at zero energy is the same as $-\mathbf{k}$, there holds (arguments \mathbf{x} and t omitted)

$$\sum_{l=0}^{\infty} \sum_{m=-l}^l f_{l,m}(0) Y_{l,m}(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{l,m}(0) Y_{l,m}(\pi - \theta, \pi + \varphi) . \quad (4.68)$$

The even terms cancel, while for the odd terms we get

$$2 \sum_{l=0}^{\infty} \sum_{m=-2l-1}^{2l+1} f_{2l+1,m}(0) Y_{2l+1,m}(\theta, \varphi) = 0 . \quad (4.69)$$

Since the spherical harmonics are linearly independent and the angles θ and φ are arbitrary, we deduce $f_{2l+1,m}(0) = 0$ for all integers $l \geq 0$ and $m = -l, \dots, l$.

- At the device boundary, one can either model ohmic contacts with a generation term [24], or prescribe a certain electron concentration n_{contact} in the form of Dirichlet boundary conditions and assume a Maxwell distribution with respect to energy. Since the zeroth order harmonic carries the charge, we thus have

$$f_{0,0}(\varepsilon) \Big|_{\text{contact}} = n_{\text{contact}} k_B T \exp\left(-\frac{\varepsilon}{k_B T}\right) . \quad (4.70)$$

In case that the generalised energy distribution function is discretised, the above expression has to be multiplied with the density of states. The other coefficients of even order harmonics are set to zero at the contact, while the coefficients of the odd order harmonics are left floating at the contact. For low-order spherical harmonics there boundary conditions have a physical interpretation: The zeroth- and second-order harmonics specify the electron concentration and the electron temperature, while the first and third order are equivalent to the electron current and the heat flux [33].

Note that the specification of two boundary conditions for the zeroth order coefficient is at first sight contradictory, since the equations for the coefficients $f_{l,m}(\mathbf{x}, \varepsilon)$ are partial differential equations of first order. However, this is resolved due to the fact that we are dealing with a *system* of partial differential equations where coefficients of odd order harmonics are unspecified everywhere along the device.

Chapter 5

Reduction of Memory Requirements

In the previous chapter we have dealt with the discretisation of the spherical harmonics expansion (SHE) equations of the Boltzmann transport equation (BTE). Implementation details have not been addressed in order to have a clear view on the calculations and manipulations which have lead to the system of linear equations (4.42). The rather complicated structure of this system makes an implementation difficult and as we shall see in this chapter, a naive programming approach can lead to memory requirements that are up to several orders of magnitude larger than actually necessary. Consequently, in this chapter the computational complexity of the discretised system (4.42) is analysed and methods are developed that reduce the overall computational complexity and memory requirements. The huge memory requirements for higher order expansions have been mentioned in a few publications [23, 24], but to the knowledge of the author of this thesis only one publication includes a short discussion on how memory requirements can be lowered [23].

5.1 The Naive Approach

When it comes to the setup of the discrete linear system (4.42) in a computer, a straightforward method is to (arbitrarily) enumerate the coefficients $f_{l,m;i}$ in the expansion (4.41), whose total number depends on the number of degrees freedom in the (x, ε) -space and the number of spherical harmonics. Then, the system matrix can be assembled as described in the previous section.

In the following, L denotes the maximum degree of the SHE and N denotes the number of degrees of freedom in the $(\mathbf{x}, \varepsilon)$ -space. We recall that there are $(L + 1)^2$ spherical harmonics of degree smaller or equal to L . It has to be emphasised that the balance equations (4.42) lead at first sight to a dense coupling of the harmonics at a given point in the $(\mathbf{x}, \varepsilon)$ -space, i.e. in the balance equation for $f_{l,m;i}$ the coefficient of $f_{l',m';i}$ is expected to be nonzero if l and l' are of different parity (cf. (3.47) and (3.51)).

In case that we do not distinguish between a coupling of unknowns due to spherical harmonics and a coupling due to the discretisation in the $(\mathbf{x}, \varepsilon)$ -space (to which we will refer to as the *naive method*), the complexity of the problem can be estimated rather easily. Since we are especially interested in the asymptotic complexity for large SHE order L and number of grid nodes N , the Landau symbol $\mathcal{O}(\cdot)$ will be used repeatedly. We also assume

that a sparse matrix format is used and that the coupling coefficients $\mathbf{v}_{l,m,l',m'}$ and $\mathbf{\Gamma}_{l,m,l',m'}$ are evaluated numerically. Thus, we assume that the coupling coefficients are nonzero for different parities of l and l' , which is justified by taking numerical noise during the numerical integration into account.

To cover the general case of unstructured grids in $(\mathbf{x}, \varepsilon)$ space, we introduce the following sparsity indicators for the use in complexity estimates later on:

Definition 10. *Given a triangulation \mathcal{T} and its dual grid \mathcal{B} with basis $P_1(\mathcal{T})$ and $P_0(\mathcal{B})$ as introduced in (4.10) and (4.11), we define the sparsity indicator by*

$$C_{\text{sparse}} := \max_{\chi \in P_0(\mathcal{B})} |\{ \varphi \in P_1(\mathcal{T}) \mid \exists \xi \in G : [\varphi(\xi)\chi(\xi) \neq 0 \vee$$

$$\exists \eta : \varphi(\xi \pm \hbar \omega_\eta)\chi(\xi) \neq 0] \}|, \quad (5.2)$$

where the notation $|A|$ denotes the number of elements of the set A and G is the simulation domain in $(\mathbf{x}, \varepsilon)$ -space. The terms $\hbar \omega_\eta$ are due to the collision processes with index η that show up in the balance equations (4.42).

We note that in the case of a structured rectangular grid in the two-dimensional (x, ε) plane and $\hbar \omega_\eta = 0$, we find $C_{\text{sparse}} = 9$, since the box around node (i, j) couples with nodes (i, j) , $(i \pm 1, j)$, $(i, j \pm 1)$ and $(i \pm 1, j \pm 1)$. We note that in case of inelastic scattering, various energy displacements $\hbar \omega_\eta$ can lead to much larger values of $C_{\text{sparse}} = 9$ even for rectangular grids.

Backed up from observations in practise, it is commonly assumed that C_{sparse} does not depend on the number of mesh elements, consequently we also assume $C_{\text{sparse}} = \mathcal{O}(1)$. With the sparsity indicator we immediately obtain:

Theorem 14. *The following statements for the resulting system matrix \mathbf{S} of (4.42) using spherical harmonics up to degree L in the naive method hold:*

- (i) \mathbf{S} is a sparse matrix of size $N(L+1)^2 \times N(L+1)^2$.
- (ii) The number of nonzero entries in each row of \mathbf{S} is at most $C_{\text{sparse}}(L+1)^2 = \mathcal{O}(L^2)$.
- (iii) The storage requirements for \mathbf{S} are $\mathcal{O}(NL^4)$.
- (iv) Matrix-vector multiplication can be realised with at most $(C_{\text{sparse}}(L+1)^2 - 1)N(L+1)^2$ additions and $C_{\text{sparse}}N(L+1)^4$ multiplications.

Proof. (i) On each of the N grid points $(L+1)^2$ unknowns are associated, hence it follows immediately that \mathbf{S} is of size $N(L+1)^2 \times N(L+1)^2$. The sparsity follows from (ii).

(ii) Consider the rows in \mathbf{S} associated with the basis function χ_i with support only on box B_i . A nonzero entry in the columns associated with φ_j requires that φ_j and χ_i have common support. There are by definition at most C_{sparse} such basis functions from $P_1(\mathcal{T})$ and since there are $(L+1)^2$ unknowns per basis function in the $(\mathbf{x}, \varepsilon)$ -space with dense coupling, there are at most $C_{\text{sparse}}(L+1)^2 = \mathcal{O}(L^2)$ nonzero entries per row.

(iii) According to (i) and (ii), \mathbf{S} consists of $N(L+1)^2$ rows with $\mathcal{O}(L^2)$ entries each. Therefore, $\mathcal{O}(NL^4)$ memory is needed to store \mathbf{S} .

(iv) From (ii) we know that there are at most $C_{\text{sparse}}(L+1)^2$ nonzero entries per row of \mathbf{S} . Thus, each entry of the result vector requires $C_{\text{sparse}}(L+1)^2 - 1$ additions and $C_{\text{sparse}}(L+1)^2$ multiplications. This effort is required for each of the $N(L+1)^2$ rows, thus we have in total $(C_{\text{sparse}}(L+1)^2 - 1)N(L+1)^2$ additions and $C_{\text{sparse}}N(L+1)^4$ multiplications. \square

Remark 1. Note that due to the parity splitting properties discussed in Section 3.2 the estimates (ii) and (iv) in Thm. 14 can be further improved by a factor of almost two, depending on L . However, since we are interested in the asymptotic behaviour as a function of L and N only, there is no need to consider this factor.

Let us briefly estimate the memory requirements for S in the case of a rectangular grid and elastic scattering only (hence $C_{\text{sparse}} = 9$) with $N = 10.000$ nodes, i.e. 100 grid points along each coordinate axis. If we choose $L = 9$, about 5×100 nonzero entries in each row have to be stored (where we already accounted for additional zeros due to parity splitting by the reduction of the sparsity indicator 9 to 5 in (ii)) for each of the 1.000.000 unknowns. We therefore have to store 500.000.000 entries, which results in 5 GB of memory assuming ten bytes per entry. However, if we select $L = 4$, only 310 MB are required. We conclude that the naive approach has huge memory requirements which more or less prohibits higher order expansions (say, $L \geq 20$) and/or extensions to higher spatial dimensions on computers available today.

Remark 2. In one spatial dimension it is for reasons of symmetry of the \mathbf{k} -space in the absence of magnetic fields sufficient to expand the distribution function into Legendre polynomials, i. e. consider only $Y_{l,0}$ with $0 \leq l \leq L$. This results in reduced storage requirements of $\mathcal{O}(NL^2)$ for the system matrix S . However, as soon as a two-dimensional spatial domain is considered, the full set of spherical harmonics is required [29], falling back to $\mathcal{O}(NL^4)$ memory requirements for S .

5.2 Sparsity of Coupling Matrices

Up to now we have assumed that the coupling among coefficients of different spherical harmonics at a given location in the $(\mathbf{x}, \varepsilon)$ -space is dense. From parity considerations in Sec. 3.2 we deduced that in fact only coefficients of spherical harmonics with different parity couple (cf. (3.47) and (3.51)). In this section we consider spherical bands and show that in this case only a sparse coupling among coefficients of different spherical harmonics occurs.

For spherical bands, the coupling is determined by the terms $\mathbf{a}_{l,m,l',m'}$ and $\mathbf{b}_{l,m,l',m'}$ as defined in (4.45) and (4.46). Before we proceed, some preparatory results are needed:

Lemma 11. *There holds*

(i)

$$\begin{aligned} \int_0^{2\pi} \cos(\varphi) \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi = \\ = \pi \times \begin{cases} (\delta_{m,m'-1} + \delta_{m,m'+1})/2, & m > 0, m' > 0, \\ \delta_{m',1}, & m = 0, m' > 0, \\ \delta_{m,1}, & m > 0, m' = 0, \\ (\delta_{m,m'-1} + \delta_{m,m'+1})/2, & m < 0, m' < 0, \\ 0, & \text{else} . \end{cases} \end{aligned} \quad (5.3)$$

(ii)

$$\begin{aligned}
\int_0^{2\pi} m \sin(\varphi) \begin{cases} -\sin(m\varphi), & m > 0 \\ 0, & m = 0 \\ \cos(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi = \\
= |m|\pi \times \begin{cases} (\delta_{|m|,|m'|-1} - \delta_{|m|,|m'|+1})/2, & m > 0, m' > 0, \\ -\delta_{m,1}, & m > 0, m' = 0, \\ (\delta_{|m|,|m'|-1} - \delta_{|m|,|m'|+1})/2, & m < 0, m' < 0, \\ 0, & \text{else.} \end{cases} \quad (5.4)
\end{aligned}$$

Proof. We start with the evaluation of

$$\int_0^{2\pi} \cos(\varphi) \cos(m\varphi) \cos(m'\varphi) d\varphi, \quad m > 0, m' > 0.$$

Using the second of the trigonometric identities

$$\begin{aligned}
\sin \theta \sin \varphi &= \frac{\cos(\theta - \varphi) - \cos(\theta + \varphi)}{2}, & \cos \theta \cos \varphi &= \frac{\cos(\theta - \varphi) + \cos(\theta + \varphi)}{2}, \\
\sin \theta \cos \varphi &= \frac{\sin(\theta + \varphi) + \sin(\theta - \varphi)}{2}, & \cos \theta \sin \varphi &= \frac{\sin(\theta + \varphi) - \sin(\theta - \varphi)}{2},
\end{aligned} \quad (5.5)$$

we obtain

$$\frac{1}{2} \int_0^{2\pi} \cos(\varphi) [\cos((m - m')\varphi) + \cos((m + m')\varphi)] d\varphi.$$

The Fourier orthogonalities ($k > 0, l \geq 0$)

$$\begin{aligned}
\int_0^{2\pi} \sin(k\varphi) \cos(l\varphi) d\varphi &= 0, \\
\int_0^{2\pi} \sin(k\varphi) \sin(l\varphi) d\varphi &= \pi \delta_{k,l}, \\
\int_0^{2\pi} \cos(k\varphi) \cos(l\varphi) d\varphi &= \pi \delta_{k,l}
\end{aligned}$$

thus yield

$$\int_0^{2\pi} \cos(\varphi) \cos(m\varphi) \cos(m'\varphi) d\varphi = \frac{\pi}{2} (\delta_{m-m',1} + \delta_{m-m',-1}).$$

Note that the symmetries

$$\cos(\theta) = \cos(-\theta), \quad \sin(\theta) = -\sin(-\theta).$$

have to be taken into account, since $m - m'$ may have arbitrary sign.

In exactly the same way the remaining expressions can be evaluated using one of the identities (5.5), the Fourier orthogonalities and taking parity into account. We skip these repetitive calculations, since they do not provide any further insight. \square

Note that the negative sign in (ii) occurs whenever $|m'| = |m| - 1$, while the positive sign appears whenever $|m'| = |m| + 1$.

Similarly, a second preparatory Lemma is necessary:

Lemma 12. *There holds*

(i)

$$\begin{aligned} \int_0^{2\pi} \sin(\varphi) \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi = \\ = \pi \times \begin{cases} (-\delta_{|m|,|m'|-1} + \delta_{|m|,|m'|+1})/2, & m > 0, m' < 0, \\ -\delta_{m',-1}, & m = 0, m' < 0, \\ -\delta_{m,-1}, & m < 0, m' = 0, \\ (\delta_{|m|,|m'|-1} - \delta_{|m|,|m'|+1})/2, & m < 0, m' > 0, \\ 0, & \text{else.} \end{cases} \end{aligned} \quad (5.6)$$

(ii)

$$\begin{aligned} \int_0^{2\pi} m \cos(\varphi) \begin{cases} -\sin(m\varphi), & m > 0 \\ 0, & m = 0 \\ \cos(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi = \\ = |m|\pi \times \begin{cases} (\delta_{|m|,|m'|-1} + \delta_{|m|,|m'|+1})/2, & m > 0, m' < 0, \\ -\delta_{m,-1}, & m < 0, m' = 0, \\ (-\delta_{|m|,|m'|-1} - \delta_{|m|,|m'|+1})/2, & m < 0, m' > 0, \\ 0, & \text{else.} \end{cases} \end{aligned} \quad (5.7)$$

Proof. This Lemma is proved in exactly the same way as the previous one, using one of the identities (5.5), the Fourier orthogonalities and taking parity into account. \square

We are now ready to prove the sparsity of the coupling term $\mathbf{a}_{l,m,l',m'}$:

Theorem 15. *For*

$$\mathbf{a}_{l,m,l',m'} = \begin{pmatrix} a_{l,m,l',m'}^1 \\ a_{l,m,l',m'}^2 \\ a_{l,m,l',m'}^3 \end{pmatrix} = \int_{\Omega} Y_{l,m} \mathbf{e}_{\varepsilon} Y_{l',m'} d\Omega,$$

as introduced in (4.45), the following statements hold:

(i)

$$a_{l,m,l',m'}^1 \neq 0 \Rightarrow l' \in \{l-1, l+1\}, \quad m' \in \{m-1, m+1\}. \quad (5.8)$$

(ii)

$$a_{l,m,l',m'}^2 \neq 0 \Rightarrow l' \in \{l-1, l+1\}, \quad m' \in \{-m-1, -m+1\}. \quad (5.9)$$

(iii)

$$a_{l,m,l',m'}^3 \neq 0 \Rightarrow l' \in \{l-1, l+1\}, \quad m' = m. \quad (5.10)$$

Proof. (i) Substitution of the definition of spherical harmonics (cf. (1.57)) into (4.45) and splitting the integral leads to

$$a_{l,m,l',m'}^1 = N_{l,m} N_{l',m'} \int_0^\pi P_l^{|m|} \sin \theta P_{l'}^{|m'|} d\theta \\ \times \int_0^{2\pi} \cos(\varphi) \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} d\varphi.$$

From Lemma 11 we deduce that $m' = m \pm 1$ is required in order to have $a_{l,m,l',m'}^1$ unequal to zero.

- $|m'| = |m| - 1$: In this case we have to consider

$$a_{l,m,l',m'}^1 = C_{l,m,l'}^- \int_0^\pi P_l^{|m|} \sin \theta P_{l'}^{|m|-1} d\theta \\ = C_{l,m,l'}^- \int_{-1}^1 P_l^{|m|} (1 - \mu^2)^{1/2} P_{l'}^{|m|-1} d\mu,$$

where

$$C_{l,m,l'}^- = \pi N_{l,|m|} N_{l',|m|-1} \times \begin{cases} 1, & m = 0 \vee m = 1, \\ \frac{1}{2}, & \text{else,} \end{cases} \quad (5.11)$$

and $N_{r,s}$ is the normalisation constant of the spherical harmonic $Y_{r,s}$ (cf. (1.57)). With (1.55) we find

$$a_{l,m,l',m'}^1 = C_{l,m,l'}^- \int_{-1}^1 \left[(l + |m| - 1) P_{l-1}^{|m|-1}(\mu) \right. \\ \left. - (l - |m| + 1) \mu P_l^{|m|-1}(\mu) \right] P_{l'}^{|m|-1} d\mu.$$

Using (1.39) for the second term, we thus obtain

$$a_{l,m,l',m'}^1 = C_{l,m,l'}^- \int_{-1}^1 \left[(l + |m| - 1) P_{l-1}^{|m|-1}(\mu) - \frac{(l - |m| + 1)}{2l + 1} \times \right. \\ \left. \left((l - |m| + 2) P_{l+1}^{|m|-1}(\mu) + (l + |m| - 1) P_{l-1}^{|m|-1}(\mu) \right) \right] P_{l'}^{|m|-1} d\mu \\ = C_{l,m,l'}^- \left[(l + |m| - 1) \delta_{l-1,l'} \right. \\ \left. - \frac{(l - |m| + 1)}{2l + 1} ((l - |m| + 2) \delta_{l+1,l'} + (l + |m| - 1) \delta_{l-1,l'}) \right],$$

Therefore, $l' = l \pm 1$ is required for nonzero $a_{l,m,l',m'}^1$.

- $|m'| = |m| + 1$: In this case we have to consider

$$a_{l,m,l',m'}^1 = C_{l,m,l'}^+ \int_0^\pi P_l^{|m|} \sin \theta P_{l'}^{|m|+1} d\theta \\ = C_{l,m,l'}^+ \int_{-1}^1 P_l^{|m|} (1 - \mu^2)^{1/2} P_{l'}^{|m|+1} d\mu,$$

where

$$C_{l,m,l'}^- = \pi N_{l,|m|} N_{l',|m|+1} \begin{cases} 1, & m = 0 \vee m = 1, \\ \frac{1}{2}, & \text{else} . \end{cases} \quad (5.12)$$

Setting $k := |m| + 1$, we have

$$\begin{aligned} a_{l,m,l',m'}^1 &= C_{l,m,l'}^+ \int_0^\pi P_l^{|m|} \sin \theta P_{l'}^{|m|+1} d\theta \\ &= C_{l,m,l'}^+ \int_{-1}^1 P_l^k (1 - \mu^2)^{1/2} P_{l'}^{k-1} d\mu, \end{aligned}$$

thus we can conclude just as in the case $|m'| = |m| - 1$ that $l' = l \pm 1$ is required for nonzero $a_{l,m,l',m'}^1$.

(ii) Substitution of the definition of spherical harmonics (cf. (1.57)) into (4.46) and splitting the integral leads to

$$\begin{aligned} a_{l,m,l',m'}^2 &= N_{l,m} N_{l',m'} \int_0^\pi P_l^{|m|} \sin \theta P_{l'}^{|m'|} d\theta \\ &\quad \times \int_0^{2\pi} \sin(\varphi) \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} d\varphi . \end{aligned}$$

From Lemma 11 we deduce that $m' = -m \pm 1$ is required in order to have $a_{l,m,l',m'}^2$ unequal to zero. Similar as in the proof of (i) we conclude that additionally $l' \in \{l-1, l+1\}$ is needed for nonzero $a_{l,m,l',m'}^2$.

(iii) Substitution of the definition of spherical harmonics (cf. (1.57)) into (4.46) and splitting the integral leads to

$$\begin{aligned} a_{l,m,l',m'}^3 &= N_{l,m} N_{l',m'} \int_0^\pi P_l^{|m|} \cos \theta P_{l'}^{|m'|} d\theta \\ &\quad \times \int_0^{2\pi} \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} d\varphi . \end{aligned}$$

The Fourier orthogonalities enforce $m' = m$, hence

$$a_{l,m,l',m'}^3 = C_{l,m,l'} \int_0^\pi P_l^{|m|} \cos \theta P_{l'}^{|m'|} d\theta \quad (5.13)$$

with

$$C_{l,m,l'} = \pi N_{l,|m|} N_{l',|m|} \begin{cases} 2, & m = 0 \\ 1, & \text{else} . \end{cases} \quad (5.14)$$

Using (1.39), we arrive at

$$\begin{aligned} a_{l,m,l',m'}^1 &= C_{l,m,l'} \int_0^\pi \frac{l - |m| + 1}{2l + 1} P_{l+1}^{|m|} P_{l'}^{|m'|} + \frac{l + |m|}{2l + 1} P_{l-1}^{|m|} P_{l'}^{|m'|} d\theta \\ &= C_{l,m,l'} \left[\frac{l - |m| + 1}{2l + 1} \delta_{l+1,l'} + \frac{l + |m|}{2l + 1} \delta_{l-1,l'} \right] . \end{aligned}$$

Thus, $l' \in \{l-1, l+1\}$ and $m' = m$ is necessary to have nonzero $a_{l,m,l',m'}^3$. \square

It would be of no big use if only $\mathbf{a}_{l,m,l',m'}$ were sparse, but $\mathbf{b}_{l,m,l',m'}$ were dense. Even though the explicit expression (4.46) does not suggest sparsity, it is nevertheless the case:

Theorem 16. *For*

$$\mathbf{b}_{l,m,l',m'} = \begin{pmatrix} b_{l,m,l',m'}^1 \\ b_{l,m,l',m'}^2 \\ b_{l,m,l',m'}^3 \end{pmatrix} = \int_{\Omega} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_{\varphi} \right) Y_{l',m'} d\Omega$$

as introduced in (4.46), the following statements for each component of $\mathbf{b}_{l,m,l',m'}$ hold:

(i)

$$b_{l,m,l',m'}^1 \neq 0 \Rightarrow l' \in \{l-1, l+1\}, \quad m' \in \{m-1, m+1\}. \quad (5.15)$$

(ii)

$$b_{l,m,l',m'}^2 \neq 0 \Rightarrow l' \in \{l-1, l+1\}, \quad m' \in \{-m-1, -m+1\}. \quad (5.16)$$

(iii)

$$b_{l,m,l',m'}^3 \neq 0 \Rightarrow l' \in \{l-1, l+1\}, \quad m' = m. \quad (5.17)$$

Proof. (i) As in the proof of the previous theorem, the first step is to substitute the definition of spherical harmonics (cf. (1.57)) into (4.46):

$$\begin{aligned} b_{l,m,l',m'}^1 &= N_{l,m} N_{l',m'} \int_{\Omega} \left[\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \cos \varphi \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \right. \\ &\quad \left. - \frac{m}{\sin \theta} P_l^{|m|}(\cos \theta) \sin(\varphi) \times \begin{cases} -\sin(m\varphi), & m > 0 \\ 0, & m = 0 \\ \cos(m\varphi), & m < 0 \end{cases} \right] \\ &\quad \times P_{l'}^{|m'|}(\cos \theta) \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} d\Omega \end{aligned}$$

Next, we rewrite the integral over the sphere Ω into integrals over the angles and separate the two integrals:

$$\begin{aligned} b_{l,m,l',m'}^1 &= N_{l,m} N_{l',m'} \left[\int_0^{\pi} \frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta P_{l'}^{|m'|}(\cos \theta) d\theta \right. \\ &\quad \times \int_0^{2\pi} \cos(\varphi) \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi \\ &\quad - \int_0^{\pi} P_l^{|m|}(\cos \theta) P_{l'}^{|m'|}(\cos \theta) d\theta \\ &\quad \times \int_0^{2\pi} m \sin(\varphi) \begin{cases} -\sin(m\varphi), & m > 0 \\ 0, & m = 0 \\ \cos(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi \left. \right] \end{aligned}$$

The integrals over the angle φ have already been evaluated in Lemma 11. It follows that for $b_{l,m,l',m'}^1$ to be nonzero, $m' \in \{m+1, m-1\}$ is required.

- $|m'| = |m| - 1$: We have

$$\begin{aligned} b_{l,m,l',m'}^1 &= C_{l,m,l'}^- \int_0^\pi \left[\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta + |m| P_l^{|m|}(\cos \theta) \right] P_{l'}^{|m|-1}(\cos \theta) d\theta \\ &= C_{l,m,l'}^- \int_{-1}^1 \left[-\frac{dP_l^{|m|}(\mu)}{d\mu} \mu(1-\mu^2)^{1/2} \right. \\ &\quad \left. + |m| P_l^{|m|}(\mu)(1-\mu^2)^{-1/2} \right] P_{l'}^{|m|-1}(\mu) d\mu \end{aligned} \quad (5.18)$$

Using (1.56) to resolve the derivative yields

$$\begin{aligned} b_{l,m,l',m'}^1 &= C_{l,m,l'}^- \int_{-1}^1 \left[l\mu^2 P_l^{|m|}(\mu) - (l+|m|)\mu P_{l-1}^{|m|}(\mu) \right. \\ &\quad \left. + |m| P_l^{|m|}(\mu) \right] P_{l'}^{|m|-1}(\mu)(1-\mu^2)^{-1/2} d\mu \end{aligned}$$

To use the orthogonality of associated Legendre functions, the term $(1-\mu^2)^{-1/2}$ has to be eliminated and the upper index of associated Legendre functions has to be equal. With (1.52) we obtain

$$\begin{aligned} b_{l,m,l',m'}^1 &= C_{l,m,l'}^- \int_{-1}^1 \left[l(l-|m|+1)\mu P_l^{|m|-1}(\mu)(1-\mu^2)^{1/2} - |m|\mu P_{l-1}^{|m|}(\mu) \right. \\ &\quad \left. + |m| P_l^{|m|}(\mu) \right] P_{l'}^{|m|-1}(\mu)(1-\mu^2)^{-1/2} d\mu \end{aligned}$$

Applying the recursions (1.39) to the first term and (1.53) to the remaining terms, we find

$$\begin{aligned} b_{l,m,l',m'}^1 &= C_{l,m,l'}^- \int_{-1}^1 \left[\frac{l(l-|m|+1)^2}{2l+1} P_{l+1}^{|m|-1}(\mu) \right. \\ &\quad \left. + \frac{l(l-|m|+1)(l+|m|)}{2l+1} P_{l-1}^{|m|-1}(\mu) \right. \\ &\quad \left. + |m|(l+|m|-1) P_{l-1}^{|m|-1}(\mu) \right] P_{l'}^{|m|-1}(\mu) d\mu \\ &= C_{l,m,l'}^- \left[\frac{l(l-|m|+1)^2}{2l+1} \delta_{l+1,l'} \right. \\ &\quad \left. + \frac{l(l-|m|+1)(l+|m|) + (2l+1)|m|(l+|m|-1)}{2l+1} \delta_{l-1,l'} \right] \end{aligned}$$

with

$$C_{l,m,l'}^- = \pi N_{l,|m|} N_{l',|m|-1} \begin{cases} 1, & m=0 \vee m=1, \\ \frac{1}{2}, & \text{else} \end{cases} \quad (5.19)$$

where $N_{r,s}$ is the normalisation constant of the spherical harmonic $Y_{r,s}$ (cf. (1.57)).

- $|m'| = |m| + 1$: Starting from

$$\begin{aligned}
b_{l,m,l',m'}^1 &= C_{l,m,l'}^- \int_0^\pi \left[\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta - |m| P_l^{|m|}(\cos \theta) \right] P_{l'}^{|m|+1}(\cos \theta) d\theta \\
&= C_{l,m,l'}^- \int_{-1}^1 \left[- \frac{dP_l^{|m|}(\mu)}{d\mu} \mu (1 - \mu^2)^{1/2} \right. \\
&\quad \left. - |m| P_l^{|m|}(\mu) (1 - \mu^2)^{-1/2} \right] P_{l'}^{|m|+1}(\mu) d\mu,
\end{aligned} \tag{5.20}$$

we arrive in analogy to the first steps in the previous case at

$$\begin{aligned}
b_{l,m,l',m'}^1 &= C_{l,m,l'}^+ \int_{-1}^1 \left[l \mu^2 P_l^{|m|}(\mu) - (l + |m|) \mu P_{l-1}^{|m|}(\mu) \right. \\
&\quad \left. - |m| P_l^{|m|}(\mu) \right] P_{l'}^{|m|+1}(\mu) (1 - \mu^2)^{-1/2} d\mu
\end{aligned}$$

With the recursion (1.54) applied to the first and the second term we find

$$\begin{aligned}
b_{l,m,l',m'}^1 &= C_{l,m,l'}^+ \int_{-1}^1 \left[\frac{l}{l + |m| + 1} (1 - \mu^2)^{1/2} \mu P_l^{|m|+1}(\mu) \right. \\
&\quad + l \frac{l - |m| + 1}{l + |m| + 1} \mu P_{l+1}^{|m|}(\mu) - (1 - \mu^2)^{1/2} P_{l-1}^{|m|+1}(\mu) \\
&\quad \left. - (l - |m|) P_l^{|m|}(\mu) - |m| P_l^{|m|}(\mu) \right] P_{l'}^{|m|+1}(\mu) (1 - \mu^2)^{-1/2} d\mu \\
&= C_{l,m,l'}^+ \int_{-1}^1 \left[\frac{l}{l + |m| + 1} (1 - \mu^2)^{1/2} \mu P_l^{|m|+1}(\mu) \right. \\
&\quad - (1 - \mu^2)^{1/2} P_{l-1}^{|m|+1}(\mu) + l \frac{l - |m| + 1}{l + |m| + 1} \mu P_{l+1}^{|m|}(\mu) \\
&\quad \left. - l P_l^{|m|}(\mu) \right] P_{l'}^{|m|+1}(\mu) (1 - \mu^2)^{-1/2} d\mu
\end{aligned}$$

The recurrence (1.39) applied to the first term and (1.55) applied to the last two terms yields

$$\begin{aligned}
b_{l,m,l',m'}^1 &= C_{l,m,l'}^+ \int_{-1}^1 \left[\frac{l}{l + |m| + 1} \frac{l - |m| + 1}{2l + 1} (1 - \mu^2)^{1/2} P_{l+1}^{|m|+1}(\mu) \right. \\
&\quad + \frac{l}{l + |m| + 1} \frac{l + |m|}{2l + 1} (1 - \mu^2)^{1/2} P_{l-1}^{|m|+1}(\mu) \\
&\quad - (1 - \mu^2)^{1/2} P_{l-1}^{|m|+1}(\mu) \\
&\quad \left. + \frac{l}{l + |m| + 1} (1 - \mu^2)^{1/2} P_{l+1}^{|m|+1}(\mu) \right] P_{l'}^{|m|+1}(\mu) (1 - \mu^2)^{-1/2} d\mu \\
&= C_{l,m,l'}^+ \left[\frac{l}{l + |m| + 1} \left(\frac{l - |m| + 1}{2l + 1} + 1 \right) \delta_{l+1,l'} \right. \\
&\quad \left. + \left(\frac{l}{l + |m| + 1} \frac{l + |m|}{2l + 1} - 1 \right) \delta_{l-1,l'} \right]
\end{aligned}$$

with

$$C_{l,m,l'}^+ = \pi N_{l,m} N_{l',m+1} \begin{cases} 1, & m = 0 \vee m = -1, \\ \frac{1}{2}, & \text{else} \end{cases},$$

where $N_{r,s}$ is the normalisation constant of the spherical harmonic $Y_{r,s}$ (cf. (1.57)).

Summing up, $l' \in \{l-1, l+1\}$ and $m' \in \{m+1, m-1\}$ is required for nonzero $b_{l,m,l',m'}^1$.

(ii) Similar to (i), the first step is to substitute the definition of spherical harmonics (cf. (1.57)) into (4.46):

$$\begin{aligned} b_{l,m,l',m'}^2 = N_{l,m} N_{l',m'} \int_{\Omega} & \left[\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \varphi \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \right. \\ & + \frac{m}{\sin \theta} P_l^{|m|}(\cos \theta) \cos(\varphi) \times \begin{cases} -\sin(m\varphi), & m > 0 \\ 0, & m = 0 \\ \cos(m\varphi), & m < 0 \end{cases} \left. \right] \\ & \times P_{l'}^{|m'|}(\cos \theta) \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} d\Omega \end{aligned}$$

Next, we rewrite the integral over the sphere Ω into integrals over the angles and separate the two integrals:

$$\begin{aligned} b_{l,m,l',m'}^2 = N_{l,m} N_{l',m'} & \left[\int_0^\pi \frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta d\theta \right. \\ & \times \int_0^{2\pi} \sin(\varphi) \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi \\ & + \int_0^\pi P_l^{|m|}(\cos \theta) d\theta \\ & \times \int_0^{2\pi} m \cos(\varphi) \begin{cases} -\sin(m\varphi), & m > 0 \\ 0, & m = 0 \\ \cos(m\varphi), & m < 0 \end{cases} \times \begin{cases} \cos(m'\varphi), & m' > 0 \\ 1, & m' = 0 \\ \sin(m'\varphi), & m' < 0 \end{cases} d\varphi \left. \right] \end{aligned}$$

The integrals in φ have already been evaluated in Lemma 12. It follows that nonzero values of $b_{l,m,l',m'}^2$ require $m' \in \{-m+1, -m-1\}$.

Unlike in the proof of (i), four instead of two cases have to be considered separately:

- $m > 0, |m'| = |m| - 1$: We obtain

$$\begin{aligned} b_{l,m,l',m'}^2 = C_{l,m,l'}^- \delta_{|m|,|m'|-1} & \int_0^\pi \left[\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta \right. \\ & \left. + |m| P_l^{|m|}(\cos \theta) \right] P_{l'}^{|m|-1}(\cos \theta) d\theta, \end{aligned}$$

which is just (5.18). Thus, nonnegative values of $b_{l,m,l',m'}^2$ require $l' \in \{l-1, l+1\}$.

- $m > 0, |m'| = |m| + 1$: We obtain

$$b_{l,m,l',m'}^2 = C_{l,m,l'}^+ \delta_{|m|,|m'|-1} \int_0^\pi \left[-\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta + |m| P_l^{|m|}(\cos \theta) \right] P_{l'}^{|m|+1}(\cos \theta) d\theta ,$$

which is (5.20) up to a factor (-1) . Thus, nonnegative values of $b_{l,m,l',m'}^2$ require $l' \in \{l-1, l+1\}$.

- $m < 0, |m'| = |m| - 1$: We obtain

$$b_{l,m,l',m'}^2 = C_{l,m,l'}^- \delta_{|m|,|m'|-1} \int_0^\pi \left[-\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta - |m| P_l^{|m|}(\cos \theta) \right] P_{l'}^{|m|-1}(\cos \theta) d\theta ,$$

which is (5.18) up to a factor (-1) . Thus, nonnegative values of $b_{l,m,l',m'}^2$ require $l' \in \{l-1, l+1\}$.

- $m < 0, |m'| = |m| + 1$: We obtain

$$b_{l,m,l',m'}^2 = C_{l,m,l'}^+ \delta_{|m|,|m'|-1} \int_0^\pi \left[\frac{dP_l^{|m|}(\cos \theta)}{d\theta} \cos \theta \sin \theta - |m| P_l^{|m|}(\cos \theta) \right] P_{l'}^{|m|+1}(\cos \theta) d\theta ,$$

which is exactly (5.20). Thus, nonnegative values of $b_{l,m,l',m'}^2$ require $l' \in \{l-1, l+1\}$.

Summing up, $l' \in \{l-1, l+1\}$ and $m' \in \{-m+1, -m-1\}$ is required for nonzero $b_{l,m,l',m'}^2$.

(iii) Substitution of the definition of spherical harmonics (cf. (1.57)) into the third component of (4.46) yields

$$b_{l,m,l',m'}^3 = N_{l,m} N_{l',m'} \int_\Omega \left[\frac{dP_l^{|m|}(\cos \theta)}{d\theta} (-\sin \theta) \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \right. \\ \left. \times P_{l'}^{|m'|}(\cos \theta) \times \begin{cases} \cos(m\varphi), & m > 0 \\ 1, & m = 0 \\ \sin(m\varphi), & m < 0 \end{cases} \right] d\Omega .$$

From the orthogonality of trigonometric functions in φ it follows that nonzero values of $b_{l,m,l',m'}^3$ induce $m = m'$. Hence, proceeding as in (i) without φ -derivative we obtain

$$b_{l,m,l',m'}^3 = C_{l,m,l'} \delta_{m,m'} \int_{-1}^1 \left[(l+|m|) P_{l-1}^{|m|}(\mu) - l\mu P_l^{|m|}(\mu) \right] P_{l'}^{|m|}(\mu) d\mu .$$

The recurrence relation (1.39) applied to the second term of the integrand leads to

$$\begin{aligned} b_{l,m,l',m'}^3 &= C_{l,m,l'} \delta_{m,m'} \int_{-1}^1 \left[(l + |m|) P_{l-1}^{|m|}(\mu) - l \frac{l - |m| + 1}{2l + 1} P_{l+1}^{|m|}(\mu) \right. \\ &\quad \left. - l \frac{l + |m|}{2l + 1} P_{l-1}^{|m|}(\mu) \right] P_{l'}^{|m|}(\mu) d\mu \\ &= C_{l,m,l'} \delta_{m,m'} \left[(l + |m|) \left(1 - \frac{l}{2l + 1} \right) \delta_{l-1,l'} - l \frac{l - |m| + 1}{2l + 1} \delta_{l+1,l'} \right]. \end{aligned}$$

The numerical constant is given by

$$C_{l,m,l'} = \pi N_{l,m} N_{l',m} \begin{cases} 2, & m = 0, \\ 1, & \text{else}, \end{cases}$$

Thus, nonzero values of $b_{l,m,l',m'}^3$ induce $l' = l \pm 1$ and $m' = m$. \square

Due to the sparsity of $\mathbf{a}_{l,m,l',m'}$ and $\mathbf{b}_{l,m,l',m'}$, Thm. 14 can be improved as follows:

Theorem 17. *In case of spherical energy bands, the following improved statements for the resulting system matrix \mathbf{S} of (4.42) using spherical harmonics up to degree L hold:*

- (i) *The number of nonzero entries in each row of \mathbf{S} is at most $9C_{\text{sparse}} = \mathcal{O}(1)$.*
- (ii) *The storage requirements for \mathbf{S} are $\mathcal{O}(NL^2)$.*
- (iii) *Matrix-vector multiplication can be realised with at most $\mathcal{O}(9C_{\text{sparse}}NL^2) = \mathcal{O}(NL^2)$ additions and multiplications.*

Proof. (i) Each unknown $f_{l,m,i}$ is coupled for fixed l and m with C_{sparse} other unknowns in $(\mathbf{x}, \varepsilon)$ -space. Additionally, from Thm. 15 and 16 we deduce that for each spatial coupling, coupling of the SHE coefficients with index $(l \pm 1, \pm m \pm 1)$ and (l, m) occurs, leading to at most $9C_{\text{sparse}}$ entries in each row.

(ii) Since each row of \mathbf{S} consists of at most $9C_{\text{sparse}}$ entries, \mathbf{S} consists of at most $9C_{\text{sparse}}N(L+1)^2 = \mathcal{O}(NL^2)$ entries.

(iii) Since there are at most $C_{\text{sparse}}N(L+1)^2 = \mathcal{O}(NL^2)$ nonzero entries in the system matrix \mathbf{S} , $\mathcal{O}(9C_{\text{sparse}}NL^2)$ additions and multiplications are needed for a matrix-vector multiplication. \square

We note again that one has to be careful with an implementation in software: In most sparse matrix implementations, an entry is allocated even if only the value zero is assigned to it. Thus, careless iteration over $\mathbf{a}_{l,m,l',m'}$ and $\mathbf{b}_{l,m,l',m'}$ for all l, m, l' and m' during matrix assembly may still lead to a sparse matrix object \mathbf{S} with memory requirements $\mathcal{O}(NL^4)$.

5.3 Matrix Factorisation

SHE can be seen as a Galerkin method in \mathbf{k} -space, whereas the discretisation (4.41) can be seen as a Galerkin method in $(\mathbf{x}, \varepsilon)$ -space for each of the coefficients of the Galerkin method

in the \mathbf{k} -space. In this section we investigate how these two Galerkin methods are linked. Consider (4.42) and assume spherical bands, thus the coefficients $\mathbf{v}_{l,m,l',m'}$ and $\mathbf{\Gamma}_{l,m,l',m'}$ are given by (4.45) and (4.46). Pulling all terms that do not depend on the integration variables out of the integral, we obtain with

$$\mathbf{a}_{l,m,l',m'} = \begin{pmatrix} a_{l,m,l',m'}^{(1)} \\ a_{l,m,l',m'}^{(2)} \\ a_{l,m,l',m'}^{(3)} \end{pmatrix}, \quad \mathbf{b}_{l,m,l',m'} = \begin{pmatrix} b_{l,m,l',m'}^{(1)} \\ b_{l,m,l',m'}^{(2)} \\ b_{l,m,l',m'}^{(3)} \end{pmatrix}, \quad \mathbf{F} = \mathbf{F}(\mathbf{x}) = \begin{pmatrix} F^{(1)}(\mathbf{x}) \\ F^{(2)}(\mathbf{x}) \\ F^{(3)}(\mathbf{x}) \end{pmatrix} \quad (5.21)$$

the balance equations (function arguments omitted whenever appropriate)

$$\begin{aligned} & \frac{\partial f_{l,m;i'}}{\partial t} \int_{B_i} H_l \varphi_{i'}(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\ & + \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} f_{l',m';i'} \int_{\partial B_i} H_l \left(\frac{\sum_{p=1}^3 F^{(p)} a_{l,m,l',m'}^{(p)}}{\mathbf{a}_{l,m,l',m'}} \right) \tilde{v}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi}) \, d\mathbf{A} \\ & - \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} f_{l',m';i'} \mathbf{b}_{l,m,l',m'} \cdot \int_{B_i} H_l \mathbf{F} \frac{\varphi_{i'}(\boldsymbol{\xi})}{\hbar \mathbf{k}(\varepsilon)} \, d\boldsymbol{\xi} \\ & = \frac{1}{Y_{0,0}} \sum_{i'=1}^N f_{0,0;i'} \sum_{\eta} \sigma_{\eta} \int_{B_i} H_l g_{l,m}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi} \pm \hbar \boldsymbol{\omega}_{\eta}) \, d\boldsymbol{\xi} \\ & \quad - \frac{1}{Y_{0,0}} \sum_{i'=1}^N f_{l,m;i'} \sum_{\eta} \sigma_{\eta} \int_{B_i} H_l \varphi_{i'}(\boldsymbol{\xi}) g_{0,0}(\varepsilon \mp \hbar \boldsymbol{\omega}_{\eta}) \, d\boldsymbol{\xi} \end{aligned} \quad (5.22)$$

where we have additionally assumed that σ_{η} is constant¹. Note that due to the assumption of spherical bands, the generalised density of states coefficients $g_{l,m}$ are identically zero for $(l, m) \neq (0, 0)$. Furthermore, we rearrange the second term as

$$\begin{aligned} & \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} f_{l',m';i'} \int_{\partial B_i} H_l \left(\frac{\sum_{p=1}^3 F^{(p)} a_{l,m,l',m'}^{(p)}}{\mathbf{a}_{l,m,l',m'}} \right) \tilde{v}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi}) \, d\mathbf{A} \\ & = \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} \sum_{p=1}^3 f_{l',m';i'} a_{l,m,l',m'}^p \int_{\partial B_i} H_l \left(\frac{F^{(p)}}{\mathbf{0}} \right) \tilde{v}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi}) \, d\mathbf{A} \\ & \quad + \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} f_{l',m';i'} \left(\frac{0}{\mathbf{a}_{l,m,l',m'}} \right) \int_{\partial B_i} H_l \tilde{v}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi}) \, d\mathbf{A} \end{aligned} \quad (5.23)$$

¹This assumption can be relaxed considerably as we shall see later on.

Thus, substituting (5.23) into (5.22) yields the balance equations

$$\begin{aligned}
& \sum_{i'=1}^V \frac{\partial f_{l,m;i'}}{\partial t} \int_{B_i} H_l \varphi_{i'}(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\
& + \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} \sum_{p=1}^3 f_{l',m';i'} a_{l,m,l',m'}^{(p)} \int_{\partial B_i} H_l \begin{pmatrix} F^{(p)} \\ \mathbf{0} \end{pmatrix} \tilde{v}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi}) \, d\mathbf{A} \\
& + \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} f_{l',m';i'} \begin{pmatrix} 0 \\ \mathbf{a}_{l,m,l',m'} \end{pmatrix} \int_{\partial B_i} H_l \tilde{v}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi}) \, d\mathbf{A} \\
& - \sum_{i'=1}^V \sum_{l'=0}^N \sum_{m'=-l'}^{l'} f_{l',m';i'} \mathbf{b}_{l,m,l',m'} \cdot \int_{B_i} H_l \mathbf{F} \frac{\varphi_{i'}(\boldsymbol{\xi})}{\hbar \tilde{k}(\varepsilon)} \, d\boldsymbol{\xi} \\
& = \frac{\delta_{0,l} \delta_{0,m}}{Y_{0,0}} \sum_{i'=1}^N f_{0,0;i'} \sum_{\eta} \sigma_{\eta} \int_{B_i} H_l g_{0,0}(\varepsilon) \varphi_{i'}(\boldsymbol{\xi} \pm \hbar \boldsymbol{\omega}_{\eta}) \, d\boldsymbol{\xi} \\
& \quad - \frac{1}{Y_{0,0}} \sum_{i'=1}^N f_{l,m;i'} \sum_{\eta} \sigma_{\eta} \int_{B_i} H_l \varphi_{i'}(\boldsymbol{\xi}) g_{0,0}(\varepsilon \mp \hbar \boldsymbol{\omega}_{\eta}) \, d\boldsymbol{\xi}
\end{aligned} \tag{5.24}$$

The crucial observation now is that there are no more terms in the integrals that depend on either l , m , l' or m' , except for H_l . However, there are only two possibilities for H_l independent of the order L of SHE: It is either unity or an exponential independent of the spherical harmonics indices. Hence, the Galerkin method in $(\boldsymbol{x}, \varepsilon)$ -space is fully decoupled from the Galerkin method in \boldsymbol{k} -space.

To account for the decoupled Galerkin methods, we propose the following enumeration strategy for the unknowns: Given a numbering of the boxes B_i with $i = 1, \dots, N$, and spherical harmonics up to degree L , we number all unknowns $f_{l,m;i}$ for fixed i within the range $[(i-1)(L+1)^2 + 1, \dots, i(L+1)^2]$. The index pair (l, m) with $l = 0, \dots, L$ and $m = -l, \dots, l$ is mapped one-to-one onto the interval $[1, (L+1)^2]$ by a function $r = r(l, m)$. There are no further restrictions to r , but typically one will choose r as simple as possible. With this enumeration of unknowns, $f_{l,m;i}$ carries the index $(i-1)(L+1)^2 + r(l, m)$.

The proposed numbering scheme induces a block structure of the system matrix \mathbf{S} . Let us introduce the matrix-valued bilinear form for the steady-state (i.e. set the time derivative

in (5.24) to zero):

$$\mathbf{a} : \mathcal{P}_1(\mathcal{T}) \times P_0(\mathcal{B}) \rightarrow \mathbb{R}^{(L+1)^2 \times (L+1)^2} \quad (5.25)$$

$$\begin{aligned} (\mathbf{a}(\varphi, \chi))_{r(l,m),r(l',m')} = & \sum_{i=1}^N \left[\sum_{p=1}^3 a_{l,m,l',m'}^{(p)} \int_{\partial B_i} H_l \begin{pmatrix} F^{(p)} \\ \mathbf{0} \end{pmatrix} \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \chi(\boldsymbol{\xi}) \, d\mathbf{A} \right. \\ & + \begin{pmatrix} 0 \\ \mathbf{a}_{l,m,l',m'} \end{pmatrix} \int_{\partial B_i} H_l \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \chi(\boldsymbol{\xi}) \, d\mathbf{A} \\ & - \mathbf{b}_{l,m,l',m'} \cdot \int_{B_i} H_l \mathbf{F} \frac{\varphi(\boldsymbol{\xi})}{\hbar \tilde{k}(\varepsilon)} \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\ & - \frac{\delta_{l,l'} \delta_{m,m'} \delta_{0,l} \delta_{0,m}}{Y_{0,0}} \sum_{\eta} \sigma_{\eta} \int_{B_i} H_l g_{0,0}(\varepsilon) \varphi(\boldsymbol{\xi} \pm \hbar \boldsymbol{\omega}_{\eta}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\ & \left. + \frac{\delta_{l,l'} \delta_{m,m'}}{Y_{0,0}} \sum_{\eta} \sigma_{\eta} \int_{B_i} H_l \varphi(\boldsymbol{\xi}) g_{0,0}(\varepsilon \mp \hbar \boldsymbol{\omega}_{\eta}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \right]. \end{aligned} \quad (5.26)$$

With this bilinear form the system matrix \mathbf{S} can be written compactly in block structure as

$$\mathbf{S} = \begin{pmatrix} \mathbf{a}(\varphi_1, \chi_1) & \mathbf{a}(\varphi_2, \chi_1) & \dots & \mathbf{a}(\varphi_N, \chi_1) \\ \mathbf{a}(\varphi_1, \chi_2) & \mathbf{a}(\varphi_2, \chi_2) & \dots & \mathbf{a}(\varphi_N, \chi_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}(\varphi_1, \chi_N) & \mathbf{a}(\varphi_2, \chi_N) & \dots & \mathbf{a}(\varphi_N, \chi_N) \end{pmatrix}, \quad (5.27)$$

which is the common matrix structure for Galerkin methods such as the finite element method. Moreover, the sparsity of \mathbf{S} can now clearly be seen: If the intersection of the support of φ_i and χ_j (taking shifts due to the scattering operator by $\pm \hbar \boldsymbol{\omega}_{\eta}$ along the energy axis into account) is empty, $\mathbf{a}(\varphi_i, \chi_j)$ is zero. Note that, in general, $\mathbf{a}(\varphi_i, \chi_j) \neq \mathbf{a}(\varphi_j, \chi_i)$ and therefore \mathbf{S} is not symmetric.

To eliminate the weak dependence on l of the integrands via H_l , we rewrite (5.25) as

$$\mathbf{a}(\varphi, \chi) = \mathbf{a}^{\text{even}}(\varphi, \chi) + \mathbf{a}^{\text{odd}}(\varphi, \chi), \quad (5.28)$$

where

$$\left(\mathbf{a}^{\text{even}}(\varphi, \chi) \right)_{r(l,m),r(l',m')} = \begin{cases} \left(\tilde{\mathbf{a}}^{\text{even}}(\varphi, \chi) \right)_{r(l,m),r(l',m')} & , \quad l \text{ even} , \\ 0 & , \quad l \text{ odd} , \end{cases} \quad (5.29)$$

and

$$\begin{aligned}
\left(\tilde{\mathbf{a}}^{\text{even}}(\varphi, \chi)\right)_{r(l,m),r(l',m')} &= \sum_{i=1}^N \left[\sum_{p=1}^3 a_{l,m,l',m'}^{(p)} \int_{\partial B_i} \begin{pmatrix} F^{(p)} \\ \mathbf{0} \end{pmatrix} \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \chi(\boldsymbol{\xi}) \, d\mathbf{A} \right. \\
&\quad + \begin{pmatrix} 0 \\ \mathbf{a}_{l,m,l',m'} \end{pmatrix} \int_{\partial B_i} \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \chi(\boldsymbol{\xi}) \, d\mathbf{A} \\
&\quad - \mathbf{b}_{l,m,l',m'} \cdot \int_{B_i} \mathbf{F} \frac{\varphi(\boldsymbol{\xi})}{\hbar \tilde{k}(\varepsilon)} \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\
&\quad - \frac{\delta_{l,l'} \delta_{m,m'} \delta_{0,l} \delta_{0,m}}{Y_{0,0}} \sum_{\eta} \sigma_{\eta} \int_{B_i} g_{0,0}(\varepsilon) \varphi(\boldsymbol{\xi} \pm \hbar \boldsymbol{\omega}_{\eta}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\
&\quad \left. + \frac{\delta_{l,l'} \delta_{m,m'}}{Y_{0,0}} \sum_{\eta} \sigma_{\eta} \int_{B_i} \varphi(\boldsymbol{\xi}) g_{0,0}(\varepsilon \mp \hbar \omega_{\eta}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \right] \quad (5.30)
\end{aligned}$$

as well as

$$\left(\mathbf{a}^{\text{odd}}(\varphi, \chi)\right)_{r(l,m),r(l',m')} = \begin{cases} 0, & l \text{ even}, \\ \left(\tilde{\mathbf{a}}^{\text{odd}}(\varphi, \chi)\right)_{r(l,m),r(l',m')}, & l \text{ odd}, \end{cases} \quad (5.31)$$

with

$$\begin{aligned}
\left(\tilde{\mathbf{a}}^{\text{odd}}(\varphi, \chi)\right)_{r(l,m),r(l',m')} &= \sum_{i=1}^N \left[\sum_{p=1}^3 a_{l,m,l',m'}^{(p)} \int_{\partial B_i} \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right) \begin{pmatrix} F^{(p)} \\ \mathbf{0} \end{pmatrix} \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \chi(\boldsymbol{\xi}) \, d\mathbf{A} \right. \\
&\quad + \begin{pmatrix} 0 \\ \mathbf{a}_{l,m,l',m'} \end{pmatrix} \int_{\partial B_i} \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right) \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \chi(\boldsymbol{\xi}) \, d\mathbf{A} \\
&\quad - \mathbf{b}_{l,m,l',m'} \cdot \int_{B_i} \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right) \mathbf{F} \frac{\varphi(\boldsymbol{\xi})}{\hbar \tilde{k}(\varepsilon)} \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\
&\quad - \frac{\delta_{l,l'} \delta_{m,m'} \delta_{0,l} \delta_{0,m}}{Y_{0,0}} \sum_{\eta} \sigma_{\eta} \int_{B_i} \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right) g_{0,0}(\varepsilon) \varphi(\boldsymbol{\xi} \pm \hbar \boldsymbol{\omega}_{\eta}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \\
&\quad \left. + \frac{\delta_{l,l'} \delta_{m,m'}}{Y_{0,0}} \sum_{\eta} \sigma_{\eta} \int_{B_i} \exp\left(\frac{\varepsilon(\mathbf{k}) + q\psi(\mathbf{x})}{k_B T}\right) \varphi(\boldsymbol{\xi}) g_{0,0}(\varepsilon \mp \hbar \omega_{\eta}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \right] . \quad (5.32)
\end{aligned}$$

On closer inspection it can now be seen that both \mathbf{a}^{even} and \mathbf{a}^{odd} are a linear combination of matrices with row and column indices $r(l, m)$ and $r(l', m')$ respectively. There are eight matrices for \mathbf{a}^{even} and eight matrices for \mathbf{a}^{odd} . Note that the matrices in \mathbf{a}^{even} have empty rows identified by odd l , while \mathbf{a}^{odd} has empty rows identified by even l . More explicitly, there holds

$$\mathbf{a}^{\text{even}}(\varphi, \chi) = \sum_{p=1}^8 \alpha_p^{\text{even}}(\varphi, \chi) \mathbf{A}_p^{\text{even}}, \quad \mathbf{a}^{\text{odd}}(\varphi, \chi) = \sum_{p=1}^8 \alpha_p^{\text{odd}}(\varphi, \chi) \mathbf{A}_p^{\text{odd}}, \quad (5.33)$$

with scalar valued functions $\alpha_p^{\text{odd}}(\varphi, \chi)$ and $\alpha_p^{\text{even}}(\varphi, \chi)$, $p = 1, \dots, 8$, the latter given by

$$\begin{aligned} \alpha_{p'}^{\text{even}}(\varphi, \chi) = & \sum_{i=1}^N \int_{\partial B_i} \begin{pmatrix} F^{(p')} \\ \mathbf{0} \end{pmatrix} \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \, d\mathbf{A} \\ & + \sum_{i=1}^N \left(\int_{\partial B_i} \tilde{v}(\varepsilon) \varphi(\boldsymbol{\xi}) \, d\mathbf{A} \right)^{(p')} \quad p' = 1, 2, 3, \end{aligned} \quad (5.34)$$

$$\alpha_{p'}^{\text{even}}(\varphi, \chi) = - \sum_{i=1}^N \int_{B_i} \mathbf{F}^{(p'-3)} \frac{\varphi(\boldsymbol{\xi})}{\hbar \tilde{k}(\varepsilon)} \, d\boldsymbol{\xi}, \quad p' = 4, 5, 6, \quad (5.35)$$

$$\alpha_7^{\text{even}}(\varphi, \chi) = - \frac{1}{Y_{0,0}} \sum_{i=1}^N \sum_{\eta} \sigma_{\eta} \int_{B_i} g_{0,0}(\varepsilon) \varphi(\boldsymbol{\xi} \pm \hbar \boldsymbol{\omega}_{\eta}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi}, \quad (5.36)$$

$$\alpha_8^{\text{even}}(\varphi, \chi) = - \frac{1}{Y_{0,0}} \sum_{i=1}^N \sum_{\eta} \sigma_{\eta} \int_{B_i} g_{0,0}(\varepsilon \mp \hbar \boldsymbol{\omega}_{\eta}) \varphi(\boldsymbol{\xi}) \chi(\boldsymbol{\xi}) \, d\boldsymbol{\xi}, \quad (5.37)$$

where the vector decomposition (5.21) was used and the notation $(\cdot)^{(p')}$ in the first expression denotes the p' -th component of the vector. The expressions for $\alpha_p^{\text{odd}}(\varphi, \chi)$ carry only the additional exponential due to the entropy term (4.37) in the integrand and thus will not be repeated here. The matrices are given by

$$(\mathbf{A}_{p'}^{\text{even/odd}})_{r(l,m),r(l',m')} = \begin{cases} a_{l,m,l',l'}^{p'} & l \text{ [even/odd]} \\ 0 & \text{otherwise} \end{cases}, \quad p' = 1, 2, 3, \quad (5.38)$$

$$(\mathbf{A}_{p'}^{\text{even/odd}})_{r(l,m),r(l',m')} = \begin{cases} b_{l,m,l',l'}^{p'-3} & l \text{ [even/odd]} \\ 0 & \text{otherwise} \end{cases}, \quad p' = 4, 5, 6, \quad (5.39)$$

$$(\mathbf{A}_7^{\text{even/odd}})_{r(l,m),r(l',m')} = \begin{cases} \delta_{l,l'} \delta_{m,m'} & l \text{ [even/odd]} \\ 0 & \text{otherwise} \end{cases}. \quad (5.40)$$

$$(\mathbf{A}_8^{\text{even/odd}})_{r(l,m),r(l',m')} = \begin{cases} \delta_{l,l'} \delta_{m,m'} \delta_{l,0} \delta_{m,0} & l \text{ [even/odd]} \\ 0 & \text{otherwise} \end{cases}. \quad (5.41)$$

Here, the notation $[a/b]$ means that either a or b can be substituted consistently within the whole expression. We note that in an implementation, $\mathbf{A}_{p'}^{\text{even}}$ and $\mathbf{A}_{p'}^{\text{odd}}$ are stored in the same matrix, since their nonzero entries are located at disjoint locations.

Summing up, the block structure (5.27) allows to write \mathbf{S} as a sum of *Kronecker products* in the form

$$\mathbf{S} = \sum_{p=1}^8 [\boldsymbol{\alpha}_p^{\text{even}} \otimes \mathbf{A}_p^{\text{even}} + \boldsymbol{\alpha}_p^{\text{odd}} \otimes \mathbf{A}_p^{\text{odd}}], \quad (5.42)$$

where \otimes denotes the Kronecker product and the coefficient matrices $\boldsymbol{\alpha}_p^{\text{even/odd}}$ are given by

$$(\boldsymbol{\alpha}_p^{\text{even/odd}})_{ij} = \alpha_p^{\text{even/odd}}(\varphi_j, \chi_i). \quad (5.43)$$

This decoupling allows another reduction of memory requirements for the matrix \mathbf{S} compared to Thm. 17:

Theorem 18. Assume spherical energy bands such that the coupling coefficients $\mathbf{v}_{l,m,l',m'}$ and $\mathbf{\Gamma}_{l,m,l',m'}$ are given by (4.45) and (4.46). Then

- (i) The system matrix \mathbf{S} can be stored with memory requirements $\mathcal{O}(N + L^2)$.
- (ii) Matrix-vector multiplication (still) requires $\mathcal{O}(NL^2)$ additions and multiplications.

Proof. (i) From (5.27) and the sparsity of the block pattern, we conclude that at most $\mathcal{O}(C_{\text{sparse}}N) = \mathcal{O}(N)$ blocks have to be stored. From (5.28) and (5.33) we see that each block $\mathbf{a}(\varphi, \chi)$ requires the storage of $16 = \mathcal{O}(1)$ coefficients. Additionally, the sparse matrices $\mathbf{A}_{p'}^{\text{even}}$ and $\mathbf{A}_{p'}^{\text{odd}}$ have to be stored for $p' = 1, \dots, 8$ independent of the degrees of freedom in $(\mathbf{x}, \varepsilon)$ -space, thus an additional memory of $\mathcal{O}(L^2)$ is required. Thus, the total memory requirements are $\mathcal{O}(N + L^2)$.

(ii) Due to the block structure (5.27), we consider a matrix-vector multiplication locally on a block first. If we decompose the vector \mathbf{x} into blocks of length $(L + 1)^2$ as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N-1} \\ \mathbf{x}_N \end{pmatrix}, \quad (5.44)$$

we obtain with (5.33)

$$\mathbf{S}\mathbf{x} = \begin{pmatrix} \sum_{s=1}^N \sum_{p=1}^8 \alpha_p^{\text{even}}(\varphi_s, \chi_1) \mathbf{A}_p^{\text{even}} \mathbf{x}_s + \alpha_p^{\text{odd}}(\varphi_s, \chi_1) \mathbf{A}_p^{\text{odd}} \mathbf{x}_s \\ \sum_{s=1}^N \sum_{p=1}^8 \alpha_p^{\text{even}}(\varphi_s, \chi_2) \mathbf{A}_p^{\text{even}} \mathbf{x}_s + \alpha_p^{\text{odd}}(\varphi_s, \chi_2) \mathbf{A}_p^{\text{odd}} \mathbf{x}_s \\ \vdots \\ \sum_{s=1}^N \sum_{p=1}^8 \alpha_p^{\text{even}}(\varphi_s, \chi_{N-1}) \mathbf{A}_p^{\text{even}} \mathbf{x}_s + \alpha_p^{\text{odd}}(\varphi_s, \chi_{N-1}) \mathbf{A}_p^{\text{odd}} \mathbf{x}_s \\ \sum_{s=1}^N \sum_{p=1}^8 \alpha_p^{\text{even}}(\varphi_s, \chi_N) \mathbf{A}_p^{\text{even}} \mathbf{x}_s + \alpha_p^{\text{odd}}(\varphi_s, \chi_N) \mathbf{A}_p^{\text{odd}} \mathbf{x}_s \end{pmatrix}. \quad (5.45)$$

Each of the $16N$ different matrix vector products $\mathbf{A}_p^{[\text{even/odd}]} \mathbf{x}_s$ has to be computed only once. Since $\mathbf{A}_{p'}^{\text{even}}$ and $\mathbf{A}_{p'}^{\text{odd}}$ can be stored together in a single matrix for each p' , only $8N$ matrix-vector products are needed, each requiring at most $\mathcal{O}(4L^2)$ additions and $\mathcal{O}(4L^2)$ multiplications. The factor of four is due to the number of nonzero entries per row, which is according to Thm. 15 and Thm. 16 at most four. These $8N$ intermediate result vectors are then linearly combined. Since \mathbf{S} is sparse, only up to $\mathcal{O}(8C_{\text{sparse}})$ additions of vectors are required per row. Before addition, $8C_{\text{sparse}}$ multiplications with either α^{even} or α^{odd} (with appropriate arguments) are required in each of the $N(L + 1)^2$ rows.

Collecting all numbers, we have in total

$$\mathcal{O}(32NL^2 + 8C_{\text{sparse}}NL^2) = \mathcal{O}(NL^2)$$

additions and multiplications. □

Some words from the implementation point of view have to be spent on the $8N$ matrix-vector products mentioned in the proof of (ii): It is in no way necessary to store all $8N$ result vectors of size $(L + 1)^2$. Instead, in an iteration over $s = 1, \dots, N$ and $p = 1, \dots, 8$, each entry of the result of the matrix vector product $(\mathbf{A}_p^{\text{even}} + \mathbf{A}_p^{\text{odd}}) \mathbf{x}_s$ can directly be

multiplied with appropriate coefficients and written to the result vector of $\mathbf{S}\mathbf{x}$. Hence, even though we have compressed the global system matrix \mathbf{S} , there is no additional memory required for a matrix vector multiplication.

Let us discuss some of the implications of Thm. 18:

- Even though the matrix can now be stored with memory requirements $\mathcal{O}(N + L^2)$, there are still $N(L + 1)^2 = \mathcal{O}(NL^2)$ unknowns in the linear system. Thus, the total memory requirements are still $\mathcal{O}(NL^2)$. However, these $N(L + 1)^2$ entries for each vector can be stored without extra overhead in a linear memory sequence, thus the proportionality constant is 1.
- Due to the block structure of \mathbf{S} , the matrix-vector multiplication can be fully parallelised, both on shared and distributed memory computers. However, further details on parallelisation are beyond the scope of this thesis.
- Even though the system matrix is stored in a compressed form, there is no extra computational effort during the assembly required. Consequently, the numerical effort for the assembly of \mathbf{S} is again $\mathcal{O}(N + L^2)$.
- It is interesting to compare the effort for matrix-vector multiplication in the compressed (Thm. 18) and in the uncompressed (Thm. 17) case: First, the coefficient of $N(L + 1)^2$ is $9C_{\text{sparse}}$ in the uncompressed case and $32 + 8C_{\text{sparse}}$ in the compressed case. For rectangles, $C_{\text{sparse}} \geq 9$, hence the compressed matrix-vector multiplication is only slightly slower than the uncompressed one. In higher spatial dimensions and/or on unstructured grids, C_{sparse} can be much larger such that the theoretical limit for the computational effort of the compressed matrix-vector multiplication is actually lower, because the prefactor of C_{sparse} is smaller. In our estimations we have not considered matrix access times for the uncompressed matrix, which further influence the resulting performance.
- For spatial dimensions smaller than three, some of the coefficients $\alpha_p^{[\text{even/odd}]}$ in (5.33) are zero due to symmetry considerations. This allows to reduce the number of matrices $\mathbf{A}_p^{[\text{even/odd}]}$ to four in one spatial dimension and six in two spatial dimensions. The constants in the complexity estimates for matrix-vector multiplication can then be improved appropriately.

To handle full-band structures, the functions $\mathbf{v}_{l,m,l',m'}(\varepsilon)$ and $\mathbf{\Gamma}_{l,m,l',m'}(\varepsilon)$ cannot be factorised as in (4.45) and (4.46), because they generally depend on the energy and on the angles θ, φ . As already suggested by Jungemann et. al. [24], a projection onto spherical harmonics of up to degree L' is convenient:

$$\mathbf{v}(\varepsilon, \theta, \varphi) \approx \sum_{l=0}^{L'} \sum_{m=-n}^n \mathbf{v}_{l,m}(\varepsilon) Y_{l,m}(\theta, \varphi) , \quad (5.46)$$

$$\mathbf{\Gamma}(\varepsilon, \theta, \varphi) \approx \sum_{l=0}^{L'} \sum_{m=-n}^n \left\{ \frac{1}{\hbar|\mathbf{k}|} \right\}_{l,m}(\varepsilon) Y_{l,m}(\theta, \varphi) , \quad (5.47)$$

where

$$\mathbf{v}_{l,m}(\varepsilon) = \int_{\Omega} \mathbf{v}(\varepsilon, \theta, \varphi) Y_{l,m}(\theta, \varphi) \, d\Omega, \quad (5.48)$$

$$\left\{ \frac{1}{\hbar|\mathbf{k}|} \right\}_{l,m}(\varepsilon) = \int_{\Omega} \frac{Y_{l,m}(\theta, \varphi)}{\hbar|\mathbf{k}(\varepsilon, \theta, \varphi)|} \, d\Omega. \quad (5.49)$$

Inserting these expansions into the expressions for $\mathbf{v}_{l,m,l',m'}$ (c.f. (4.43)) and $\mathbf{\Gamma}_{l,m,l',m'}$ (c.f. (4.44)), we obtain

$$\mathbf{v}_{l,m,l',m'}(\varepsilon) = \sum_{l''=0}^{L'} \sum_{m''=-l''}^{l''} \mathbf{v}_{l'',m''}(\varepsilon) \mathbf{a}_{l,m,l',m',l'',m'',} \quad (5.50)$$

$$\mathbf{\Gamma}_{l,m,l',m'}(\varepsilon) = \sum_{l''=0}^{L'} \sum_{m''=-l''}^{l''} \left\{ \frac{1}{\hbar|\mathbf{k}|} \right\}_{l'',m''}(\varepsilon) \mathbf{b}_{l,m,l',m',l'',m''} \quad (5.51)$$

where

$$\mathbf{a}_{l,m,l',m',l'',m''} = \int_{\Omega} Y_{l,m} Y_{l',m'} Y_{l'',m''} \, d\Omega \quad (5.52)$$

$$\mathbf{b}_{l,m,l',m',l'',m''} = \int_{\Omega} \left(\frac{\partial Y_{l,m}}{\partial \theta} \mathbf{e}_{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_{l,m}}{\partial \varphi} \mathbf{e}_{\varphi} \right) Y_{l',m'} Y_{l'',m''} \, d\Omega. \quad (5.53)$$

Note that in case that $L' = 0$ is sufficient to describe the velocity vector $\mathbf{v}(\varepsilon, \theta, \varphi)$ and the wave vector $\mathbf{k}(\varepsilon, \theta, \varphi)$, i.e. $|\mathbf{v}(\varepsilon, \theta, \varphi)| = \tilde{v}(\varepsilon)$ and $|\mathbf{k}(\varepsilon, \theta, \varphi)| = \tilde{k}(\varepsilon)$, we fall back to (4.45) and (4.46) as expected.

Theorem 19. *With an approximation of $\mathbf{v}(\varepsilon, \theta, \varphi)$ and $\mathbf{\Gamma}(\varepsilon, \theta, \varphi)$ by a SHE up to degree L' as given in (5.46) and (5.47), there holds*

- (i) *The memory required for the storage of the system matrix \mathbf{S} is at most $\mathcal{O}(N(L')^2 + (L')^2 L^4)$.*
- (ii) *The matrix-vector product then requires $\mathcal{O}(N(L')^2 L^4)$ operations.*

Thus, it is possible to achieve a memory reduction by a factor $\mathcal{O}(L^4/(L')^2)$ (typically $N(L')^2 \gg (L')^2 L^4$ holds) even in the case of fullband structures compared to the naive scheme of storing each nonzero entry separately. However, in contrast to Thm. 18 for spherical bands, this reduction comes at the expense of increased computational effort for matrix-vector multiplication compared to the naive scheme analysed in Thm. 14.

Proof. A rigorous proof of Thm. 19 requires the generalisation of the steps that have lead to Thm. 18. These generalisations mainly consist of an additional summation over l'' and m'' , so that we only outline the necessary steps here:

(i) Substitution of (5.50) and (5.51) into (4.42) and repeating the steps that have lead to the bilinear form $\mathbf{a}(\cdot, \cdot)$ in (5.26), then splitting into contributions from even and odd spherical harmonics as in (5.28), one obtains that $\mathbf{a}^{\text{even}}(\varphi, \chi)$ and $\mathbf{a}^{\text{odd}}(\varphi, \chi)$ can be written as a sum of $6(L' + 1)^2 + 2$ matrices $\mathbf{A}_p^{\text{[even/odd]}}$ each. These matrices require at most $(6(L' + 1)^2 + 2)(L + 1)^4 = \mathcal{O}((L')^2 L^4)$ memory in total. Presumably it is possible to prove

a sparsity pattern for each of these $\mathbf{A}_p^{[\text{even/odd}]}$, but for the moment we assume that at each of them is densely populated.

Additionally, the storage are $12(L' + 1)^2 + 2$ coefficients $\alpha_p^{\text{even}}(\varphi, \chi)$ and $\alpha_p^{\text{odd}}(\varphi, \chi)$ needed, resulting in the need for $\mathcal{O}(N(L')^2)$ memory. Therefore, in total $\mathcal{O}(N(L')^2 + (L')^2 L^4)$ memory is needed.

(ii) The costs of matrix-vector multiplication are obtained as in Thm. 18, taking into account that there are now $6(L' + 1)^2 + 2$ block matrices $\mathbf{A}_p^{[\text{even/odd}]}$ instead of eight: One entry of the result vector of the matrix-vector product now requires $6(L' + 1)^2 + 2$ vector products with $\mathcal{O}(L^2)$ operations each, hence the computation of all $N(L + 1)^2$ entries of the result vector require $\mathcal{O}(N(L')^2 L^4)$ operations. Again, if there is a sparsity pattern for each of the matrices $\mathbf{A}_p^{[\text{even/odd}]}$, the estimate improves to $\mathcal{O}(N(L')^2 L^2)$ operations. \square

Chapter 6

Implementation and Results

In this chapter we discuss the results obtained from the numerical solution of the spherical harmonics expansion (SHE) equations derived from the Boltzmann transport equation (BTE) in the preceding chapters. However, due to the hyperbolic nature of the underlying continuous system of first-order partial differential equations, several numerical challenges have to be passed. We first discuss the main problems that inherently show up due to the mathematical structure of the equations. After that we give all the parameters used for the simulations and point at pitfalls that have shown up during the development of the program code. Finally, the results for a n^+nn^+ diode are discussed.

6.1 Problem Setup

In this thesis a n^+nn^+ diode as shown in Fig. 6.1 is simulated. Even though such a device is of no use in practise, it serves as model for a MOSFET in inversion: The highly doped source and drain contacts are then identified with the n^+ -regions in the diode, while the channel in inversion corresponds to the n -region with lower doping. This simplification, however, does not allow for a full characterisation of the MOSFET, but provides a qualitative analysis of quantities such as carrier velocities in the channel.

The diode is divided into three parts of equal size. The left and the right parts are the n^+ regions, while the lower doped n region is located in the centre. The length of the diode is 200nm. The voltage at the right contact was set to 0.3, 1.0 and 3.0 Volt, while the left contact is grounded. The band structure was assumed to be parabolic, thus the following relationships hold:

$$\varepsilon = \frac{\hbar^2 |\mathbf{k}|^2}{2m^*} \quad (6.1)$$

$$\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}} \varepsilon = \frac{\hbar \mathbf{k}}{m^*} = \sqrt{2\varepsilon/m^*} \quad (6.2)$$

$$\hbar |\mathbf{k}| = \sqrt{2m^* \varepsilon} \quad (6.3)$$

The scattering operator is assumed to consist of acoustic and optical phonon scattering [27, 33]. The former is assumed to be isotropic and completely elastic, thus the scattering term is given by

$$s_{ac}(\mathbf{x}, \mathbf{k}, \mathbf{k}') = \sigma_{ac} \delta(\varepsilon(\mathbf{k}') - \varepsilon(\mathbf{k})) , \text{ where } \sigma_{ac} = \frac{2\pi k_B T}{\hbar V u_l^2 \rho} D_1^2 . \quad (6.4)$$

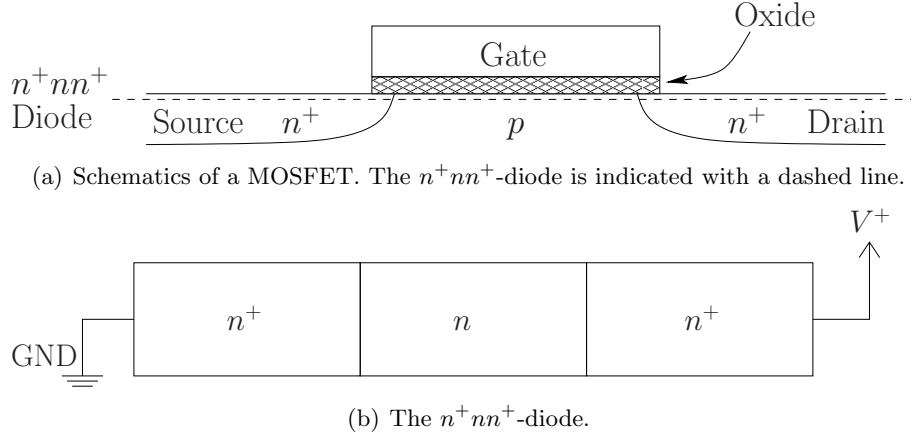


Figure 6.1: The investigation of an n^+nn^+ -diode is motivated from the channel of a MOSFET in inversion, where the p -region close the the gate oxide accumulates electrons and thus becomes a thin n -channel.

The quantity V is the volume, which cancels after appropriate scaling. The other quantities are explained in Tab. 6.1. Optical phonon scattering is modelled by the inelastic term

$$s_{\text{op}}(\mathbf{x}, \mathbf{k}, \mathbf{k}') = \frac{\pi(D_t K)^2}{\rho\omega_{\text{op}}} \times \begin{cases} N_{\text{op}}\delta(\varepsilon(\mathbf{k}') - \varepsilon(\mathbf{k}) - \hbar\omega_{\text{op}}), & \text{(absorption),} \\ (N_{\text{op}} + 1)\delta(\varepsilon(\mathbf{k}') - \varepsilon(\mathbf{k}) + \hbar\omega_{\text{op}}), & \text{(emission),} \end{cases} \quad (6.5)$$

where the number of optical phonons N_{op} is given by Bose-Einstein statistics

$$N_{\text{op}} = \frac{1}{\exp(\frac{\hbar\omega_{\text{op}}}{k_B T}) - 1}. \quad (6.6)$$

Note that the above scattering terms have to be transformed from momentum-space (\mathbf{k} -space) to energy space, which adds a multiplication with the density of states to the scattering rates (cf. (2.65)). We use these two scattering mechanisms only, others are described in the literature and can be included in the context of SHE similarly [24, 27, 29, 33, 48].

6.2 Pitfalls

During implementation and testing of the simulator, several issues had to be dealt with. Such pitfalls are often mentioned – if at all – only in a single sentence in publications, even though a lot of time was spent on dealing with them. In this section we discuss the pitfalls observed.

6.2.1 Initial Guess for the Potential

A glimpse on the SHE equation (3.30) reveals that the force term \mathbf{F} and thus the potential within the device must be given. Since it is very hard or even impossible to write the potential even in a simple one-dimensional n^+nn^+ -diode in closed form, we compute the potential ψ in the device by numerical solution of the drift-diffusion model in its simplest

Parameter	Symbol	Value
Elementary charge	$ q $	$1.6022 \times 10^{-19} \text{ C}$
Planck constant	\hbar	$6.626 \times 10^{-34} \text{ Js}$
Boltzmann constant	k_B	$1.381 \times 10^{-23} \text{ J/K}$
Lattice temperature	T	300 K
Sound speed	u_l	9000 m/s
Mass density	ρ	2330 kg m^{-3}
Effective mass	m^*	$0.26 \times m_0$
Electron rest mass	m_0	$9.1094 \times 10^{-31} \text{ kg}$
Lattice deformation potential	D_l	9.00 eV
Optical phonon coupling constant	$D_t K$	500 eV/m
Optical phonon energy	$\hbar\omega_{\text{op}}$	0.05 eV

Table 6.1: Input parameters for the simulation of an n^+nn^+ diode. Note that 1 eV equals 1.6022×10^{-19} Joules.

form:

$$-\Delta\psi = \frac{|q|}{\epsilon}(p - n - N_D) , \quad (6.7)$$

$$\nabla \cdot (\nabla n - V_T n \nabla \psi) = 0 , \quad (6.8)$$

$$\nabla \cdot (\nabla p + V_T p \nabla \psi) = 0 , \quad (6.9)$$

where the *thermal voltage* $V_T = k_B T / |q|$ is approximately 26mV at room temperature, N_D is the net doping concentration and n and p are the unknown electron and hole concentrations respectively. The dielectric constant ϵ is not to be confused with the energy ϵ . This minimalistic set of equations neglects many additional physical effects, but provides a sufficiently accurate initial guess for the potential. For numerical stabilisation the Scharfetter-Gummel scheme as described in Sec. 4.2 was used. Due to the exponential dependence of the carrier concentrations on the potential, a direct linearisation of the system typically fails to converge. Thus, we rely on the so-called *Gummel scheme* (see e.g. [42]), where we successively solve the three equations, but use the relations

$$n = n_i \exp\left(\frac{\psi - \varphi_n}{V_T}\right) , \quad p = n_i \exp\left(\frac{\varphi_p - \psi}{V_T}\right) \quad (6.10)$$

in the Poisson equation (6.7). The linearised equation for the update $\delta\psi^k$ at the k -th iteration step with $\lambda := |q|/\epsilon$ then is

$$-\Delta(\delta\psi^k) + \lambda \frac{n^k + p^k}{V_T}(\delta\psi^k) = \Delta\psi^k + \lambda(p^k - n^k - N_D) . \quad (6.11)$$

For further details on the numerical solution of the drift diffusion model we refer to the literature, e.g. [42].

6.2.2 Degenerate or Decoupled System

Let us consider a first-order SHE of the BTE for spherical bands in one spatial dimension aligned along the z -direction. Due to the symmetry of the problem we can conclude that the coefficient of $Y_{l,m}$ is zero in case that $m \neq 0$. The coupling matrices (4.45) and (4.46) can then be computed as

$$\mathbf{a}_{l,m,l',m'} = \begin{pmatrix} 0 & 1/\sqrt{3} \\ 1/\sqrt{3} & 0 \end{pmatrix}, \quad (6.12)$$

$$\mathbf{b}_{l,m,l',m'} = \begin{pmatrix} 0 & 0 \\ 2/\sqrt{3} & 0 \end{pmatrix}, \quad (6.13)$$

so that the resulting system of equations in steady state becomes

$$\begin{aligned} \tilde{v}(\varepsilon)/\sqrt{3} \left(\frac{\partial f_1}{\partial z} + F \frac{\partial f_1}{\partial \varepsilon} \right) &= Q_0\{f\}, \\ \tilde{v}(\varepsilon)/\sqrt{3} \left(\frac{\partial f_0}{\partial z} + F \frac{\partial f_0}{\partial \varepsilon} \right) - \frac{2F}{\sqrt{3}\hbar\tilde{k}(\varepsilon)} \frac{\partial f_0}{\partial \varepsilon} &= Q_1\{f\}, \end{aligned} \quad (6.14)$$

where $\tilde{v}(\varepsilon)$ and $1/(\hbar\tilde{k})$ stem from (4.45) and (4.46). We observe the following:

- In a parabolic band approximation, $|\mathbf{k}| \sim \sqrt{\varepsilon}$, thus the coefficient $2F/(\sqrt{3}\hbar\tilde{k}(\varepsilon))$ is singular at zero energy. Consequently, extra care has to be taken there: Since the singularity is integrable, a higher order numerical quadrature should be sufficient. Additionally, there are boundary-conditions at zero energy which allow to circumvent these problems.
- If the force F is zero at some point, then all energy derivatives vanish. Conversely, if the force is very large, the prefactor of the z -derivative can become very small compared to the prefactor of the energy derivatives.
- The equations are coupled via the collision operator Q only. Thus, if $Q \equiv 0$, the system decouples into two separate equations. Such a decoupling is not desired: Since boundary conditions are defined for the even harmonics only (cf. Sec. 4.5), the decoupled equation for the odd harmonics would then be underdetermined, while the equation for the even harmonics would be overdetermined due to the presence of two boundary conditions for a first order partial differential equation. Thus, it is not possible to investigate the numerical behaviour of the left hand side of (6.14) by setting $Q \equiv 0$, which complicates any analysis considerably.

6.2.3 Linear Solver and Preconditioner

Let us consider a uniformly doped semiconductor in equilibrium. The distribution function is then given by a Maxwellian:

$$f(x, \varepsilon) = n_{\text{doping}} k_B T \exp\left(-\frac{\varepsilon}{k_B T}\right), \quad (6.15)$$

where n_{doping} is the concentration of dopants. We denote with $\mathcal{M}(\varepsilon) := \exp(-\varepsilon/(k_B T))$ the exponential term. Choosing the simulation domain to range from zero to one electron

volt, evaluation of \mathcal{M} at the extremal values yields

$$\mathcal{M}(0 \text{ eV}) = 1, \quad \mathcal{M}(1 \text{ eV}) \approx 6.2 \times 10^{-16}. \quad (6.16)$$

While such a large variation in magnitude does not pose any analytical problems, it certainly does in computer with finite precision floating point units. For example, the IEEE Standard for Floating-Point Arithmetic (IEEE 754) specifies a double precision format of 64 bit with a mantissa length of 53 bits, thus a relative accuracy of about 16 decimal digits can be obtained. However, since the spherical harmonics coefficients have values of similar magnitude when evaluated at neighbouring grid points, the finite precision arithmetic is sufficient to perform a matrix-vector multiplication of good accuracy with the system matrix.

Consider now the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ with initial guess \mathbf{x}_0 and initial residual $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$. If this system is solved using a Krylov subspace method, a solution is sought in the m -dimensional subspace

$$\mathcal{K}_m = \text{span} \{ \mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \mathbf{A}^2\mathbf{r}_0, \dots, \mathbf{A}^{m-1}\mathbf{r}_0 \}. \quad (6.17)$$

The next iterate \mathbf{x}_m is then chosen from the affine subspace $\mathbf{x}_0 + \mathcal{K}_m$, where different Krylov methods lead to different choices of \mathbf{x}_m . For demonstration purposes, let us continue with the Generalised Minimum Residual Method (GMRES) [40], where \mathbf{x}_m is chosen to be the vector that minimises the residual among all vectors from $\mathbf{x}_0 + \mathcal{K}_m$. However, since small relative deviations of the approximate solution f_h near zero energy lead to variations in the residual of, say, δ , the same relative variations at energies of 1eV lead to variations in the residual of size $\delta \times 10^{-16}$ only. Thus, the numerical solution f_h will suffer from increased numerical noise at higher energies, because deviations from the true solution at zero energy have a much higher weight than deviations at high energies.

A method to reduce these oscillations is the following: Given an unknown $f_{n,m;i}$ located at location $(\mathbf{x}_i, \varepsilon_i)$, we expect due to reasons of physics at equilibrium that $f_{n,m;i}$ is by a factor of approximately $\exp(\varepsilon_i/(k_B T))$ smaller than the unknown $f_{n,m;i'}$ located at $(\mathbf{x}_i, 0)$. Thus, we replace

$$f_{n,m;i} \mapsto f_{n,m;i}^{\text{exp}} = f_{n,m;i} \exp\left(\frac{\varepsilon_i}{k_B T}\right) \quad (6.18)$$

and repeat this procedure for all unknowns. In case that the true distribution function is close the equilibrium given by (6.15), the discrete solution is then given by

$$f_{n,m;i}^{\text{exp}} \approx \alpha_{n,m} \quad \forall i, \quad (6.19)$$

where $\alpha_{n,m}$ depends only on the indices n and m of the respective spherical harmonic. In a purely mathematical view, the rescaling of unknowns corresponds to an application of a left- and a right-preconditioner \mathbf{P} : If we write the unscaled system as $\mathbf{A}\mathbf{x} = \mathbf{b}$, denote the scaled vector of unknowns with \mathbf{x}' and the identity matrix with \mathbf{I} , then the substitution (6.18) is equivalent to

$$\mathbf{A}\mathbf{x} = \mathbf{b} \mapsto \mathbf{A}\mathbf{I}\mathbf{x} = \mathbf{b} \quad (6.20)$$

$$\Leftrightarrow \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{x} = \mathbf{b} \quad (6.21)$$

$$\Leftrightarrow \mathbf{A}\mathbf{P}^{-1}\mathbf{x}' = \mathbf{b}, \quad (6.22)$$

where

$$\mathbf{P} = \begin{pmatrix} \exp(\varepsilon_0/(k_B T)) & 0 & \dots & \dots & 0 \\ 0 & \exp(\varepsilon_1/(k_B T)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & \exp(\varepsilon_{N-1}/(k_B T)) \end{pmatrix} \in \mathbb{R}^{N \times N}. \quad (6.23)$$

Since the boundary conditions at the device contacts are of Dirichlet type and given by a Maxwellian, the right hand side vector \mathbf{b} also consists of entries of different orders of magnitude. Additionally, since \mathbf{A} has entries varying by a few orders of magnitude only, the rescaled matrix $\mathbf{A}\mathbf{P}^{-1}$ consists of entries with a much larger variation in magnitude. This issue can be resolved by a left-multiplication with \mathbf{P} , which reverses the scaling induced by \mathbf{P}^{-1} and additionally leads to right hand side entries of similar magnitude:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{A}\mathbf{P}^{-1}\mathbf{x}' = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{P}\mathbf{A}\mathbf{P}^{-1}\mathbf{x}' = \mathbf{P}\mathbf{b}. \quad (6.24)$$

This combined left- and right-preconditioner, which we refer to as *exponential preconditioner*, was tested and compared with two other basic preconditioners:

- (i) A diagonal left-preconditioner \mathbf{D} , which scales the diagonal of \mathbf{A} to consist of ones only, thus

$$\mathbf{D} = \begin{pmatrix} a_{0,0}^{-1} & 0 & \dots & \dots & 0 \\ 0 & a_{1,1}^{-1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & a_{N-1,N-1}^{-1} \end{pmatrix} \in \mathbb{R}^{N \times N}. \quad (6.25)$$

Here, $\delta_{i,j}$ denotes the Kronecker delta. We will refer to this preconditioner as *diagonal preconditioner*.

- (ii) A diagonal left-preconditioner \mathbf{R} , which scales the largest element in modulus of each row to one, thus

$$\mathbf{R} = \begin{pmatrix} (\max_j |a_{0,j}|)^{-1} & 0 & \dots & \dots & 0 \\ 0 & (\max_j |a_{1,j}|)^{-1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & (\max_j |a_{N-1,j}|)^{-1} \end{pmatrix} \in \mathbb{R}^{N \times N}. \quad (6.26)$$

Again, $\delta_{i,j}$ denotes the Kronecker delta. We will refer to this preconditioner as *row scaling preconditioner*.

The comparison of preconditioners shown in Tab. 6.2 shows some interesting results: Most notably, the condition number of the system matrix without preconditioner is too large for double precision arithmetic, thus it is required to use either high-precision arithmetic or to use a preconditioner. Surprisingly, the diagonal preconditioner leads to an even higher condition number of about 10^{19} . However, it has been observed that the diagonal preconditioner performs better if lower scattering rates are chosen. In contrast, the row scaling preconditioner reduces the condition number by several order of magnitude to 10^7 and

DOF/axis	No precondition.	Diagonal	Row scaling	Exponential
30	2.1914×10^{17}	1.8352×10^{19}	9.9286×10^6	1.9107×10^{11}

Table 6.2: Comparison of condition numbers with different preconditioners for the system matrix resulting from the simulation of a n^+nn^+ -diode on a uniform grid with a SHE of first order.

performed best even at lower scattering rates. Interestingly, the exponential preconditioner does not perform as well and leads to condition numbers in the range 10^{11} , even though the unknowns are scaled such that they are of similar order of magnitude. Nevertheless, the exponential preconditioner yields better results than the diagonal preconditioner.

The action of the preconditioners on the eigenvalue spectrum of the system matrix is shown in Fig. 6.2. The row scaling and the exponential preconditioners transform the eigenvalues of the original matrix \mathbf{A} to a cluster of eigenvalues with both real and imaginary part between -1 and 1 . The diagonal preconditioner also clusters most eigenvalues in that interval except for a few large eigenvalues that lead to high condition numbers. These few large eigenvalues most likely stem from the discretisation near zero energy, where diagonal entries are smaller than the off-diagonal entries. Interestingly, even though the exponential preconditioner possesses an eigenvalue spectrum similar to the row scaling preconditioner, the matrix condition numbers differ by several orders of magnitude. Hence, we can conclude that the antisymmetric part of the system matrix is not just a small perturbation of the system matrix.

Another important thing to note in Fig. 6.2 is that eigenvalues with both positive and negative sign occur. The unpleasant consequence is that many iterative solvers like the conjugate gradient method cannot be used and more general variants like GMRES have to be used. The alternative is to rely on direct solvers, which do not allow for very large numbers of unknowns. The solver package PETSc [2] was used to benchmark several different iterative solvers. The result for low scattering rates is disappointing: None of the iterative solver was able to solve the system matrix arising from 25 grid points in each coordinate direction, no matter which preconditioner was chosen. This is especially surprising since the matrix condition number is moderate if row scaling is used. For this reason, a direct Gauss solver with pivoting was used in the following. Since simulations with higher scattering rates have been carried out only after this change of the linear solver, iterative solvers have not been tested with higher scattering rates.

Another improvement used by Jungemann et. al. [24] is to eliminate all unknown expansion coefficients of odd order in a preprocessing step. Such an elimination is numerically cheap due to the results from Chapter 5, since odd order coefficients couple with even order coefficients only. Such a reduction reduces the number of unknowns by a factor of two in the case of one-dimensional devices, while the benefit is even higher in higher dimensions. Consequently, the resulting system matrix is smaller and a direct solver is by a factor of eight faster¹. In Fig. 6.3 the action of the elimination process on the eigenvalues of the system matrix is depicted. Eigenvalues larger than 0.5 do not appear anymore (apart from the eigenvalues 1 due to boundary conditions) and the remaining eigenvalue spectrum remains

¹Under the assumption that for a matrix of size $N \times N$ the computational effort for the solution process is $\mathcal{O}(N^3)$

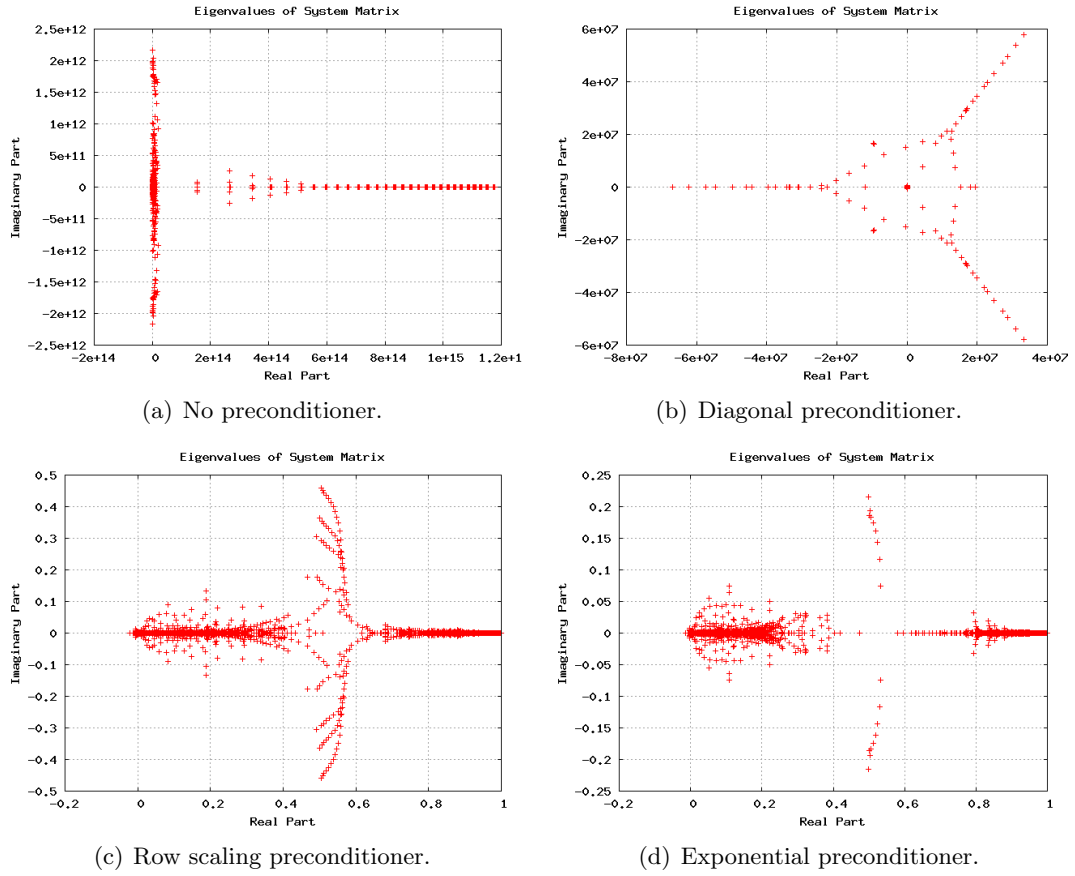


Figure 6.2: Eigenvalues of the system matrix resulting from the simulation of an n^+nn^+ -diode with different preconditioners. Note that due to the inclusion of boundary conditions in the matrix there is an eigenvalue 1 in all four cases.

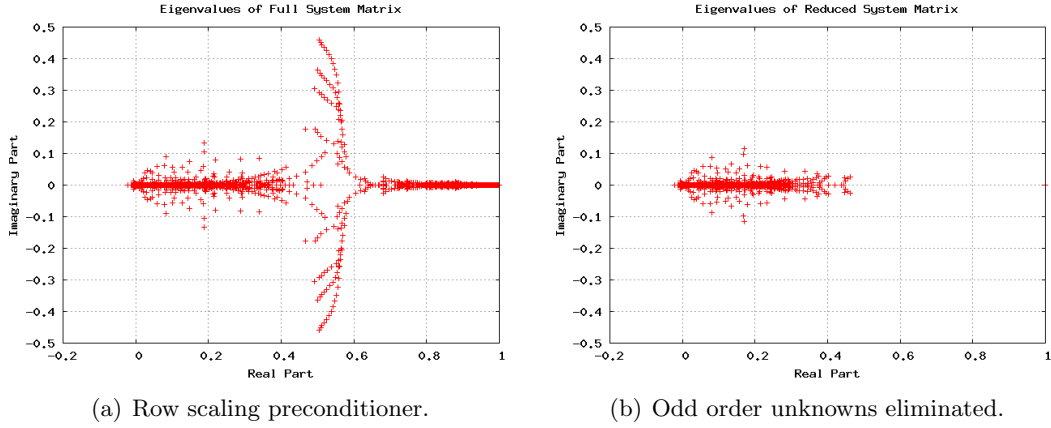


Figure 6.3: Comparison of eigenvalues of the system matrix before and after elimination of odd order unknowns. Prior to the elimination step, a row scaling preconditioner has been applied.

almost unchanged. The condition number decreases by another factor of approximately five. However, there are still several negative eigenvalues present, thus excluding many iterative solvers that require a positive definite matrix.

6.3 Results for a n^+nn^+ Diode

Even though physical quantities have been written explicitly throughout this thesis, one must not expect highly accurate results in terms of physics at this point, because the simulation results are not fitted to measurement data from a real device. In particular, since a thorough investigation and quantitative comparison with existing models was beyond the scope of this thesis, one simply cannot expect that the obtained simulation results provide an accurate description of a real device. In particular, a consideration of acoustical and optical phonon scattering only turned out to yield unphysical small scattering rates in total. Therefore, the acoustical phonon scattering rate (6.4) was increased by a factor of 100 in order to obtain physically more meaningful results. Nevertheless, one has to bear in mind that a lot more parameter tweaking is needed for reliable and predictive simulation results.

The expansion coefficients $f_{0,0}$, $f_{1,0}$ of the generalised energy distribution function \tilde{f} as well as the computed electron densities and the resulting potential for a n^+nn^+ diode of 200nm length and SHE up to first order are shown and compared with the results obtained from a simplistic drift-diffusion model (DDM) in Fig. 6.4. As can be seen from the grid, the mesh was refined near the transition regions where the electric field is large, whereas the grid spacing in energy direction is uniform. Both the electron concentration as well as the potential inside the device computed by DDM and SHE are very similar. In fact, this is not surprising, since only a low external field is applied and the internal fields due to differences in doping concentrations are small. The distribution function does not have any spurious oscillations and one can see that at $x \approx 130\text{nm}$ a local maximum of the distribution function coefficients with respect to the spatial variable x develops. In a physical interpretation this means that the electrons are accelerated in the lower doped centre region and thus gain energy. As soon as they hit the higher doped n^+ -region on the right, scattering dominates

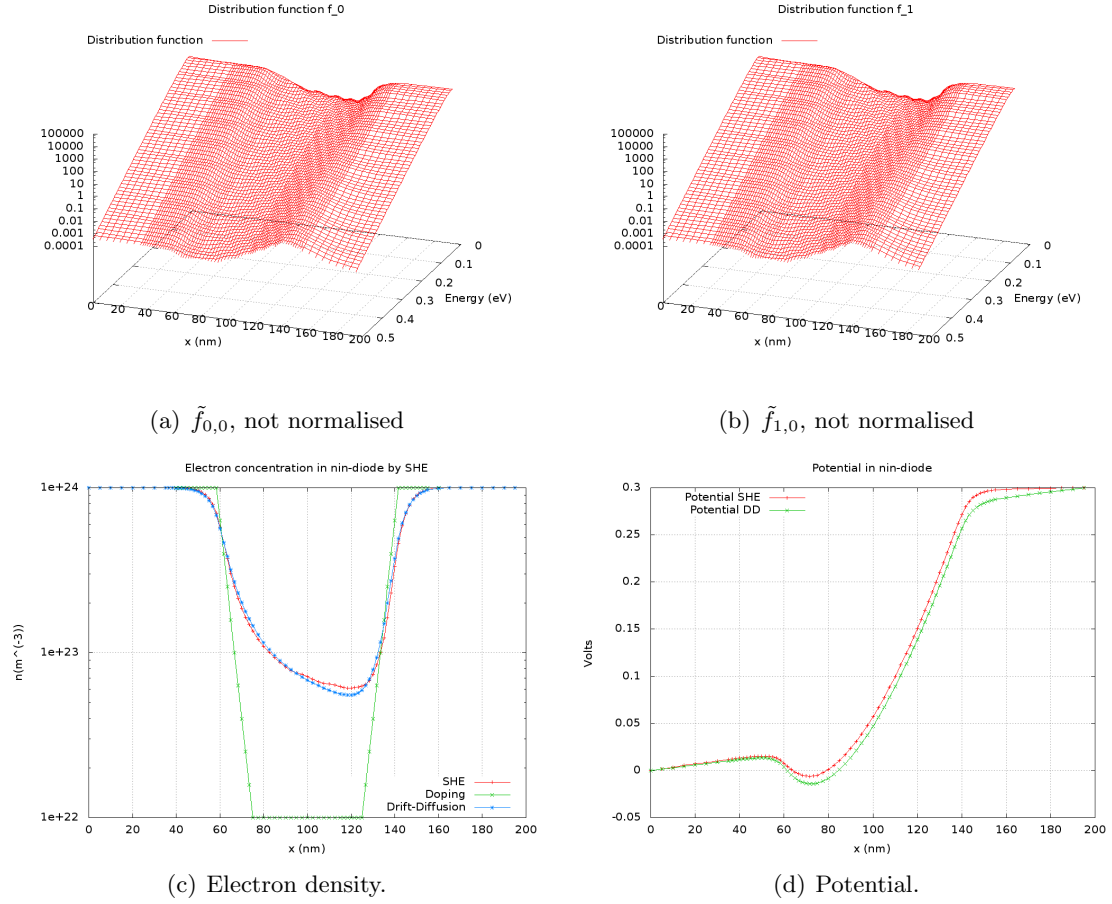


Figure 6.4: Self-consistent solution of SHE equations at an applied voltage of 0.3 V.

such that electrons lose energy. Near the contact the distribution function with respect to energy is already close to a Maxwell-distribution again.

At higher applied fields the situation the so-called *high energy tail* of the distribution function can readily be seen around $x \approx 130\text{nm}$ in Fig. 6.5. Moreover, the plot for the zeroth-order coefficient $f_{0,0}$ shows that the choice of Dirichlet boundary data is not the best: At higher energies the distribution function near the contact at $x = 200\text{nm}$ is too far away from the equilibrium Maxwell distribution, thus there are jumps in the solution. Thus, a generation/recombination term at the contact as discussed in Sec. 4.5 is of advantage. Moreover, the numerical solution tends to produce spurious oscillations near the contact in such cases. The oscillations in the coefficients of the distribution function near $x = 60\text{nm}$ in Fig. 6.5 are due to artefacts from the energy boundary at $\varepsilon = \varepsilon_{\max}$ and do not influence the computed macroscopic quantities.

The norm of the potential update vector at step s , $\|\Delta\psi_s\|$, of the Gummel iteration scheme for the solution of the coupled Boltzmann-Poisson system is shown in Fig. 6.6. While convergence is fast for small voltages and a small number of iteration steps suffice, convergence is very slow for higher voltages of several Volts. The same is true for the Gummel iteration used in the drift-diffusion model which serves as a first initial guess for the potential.

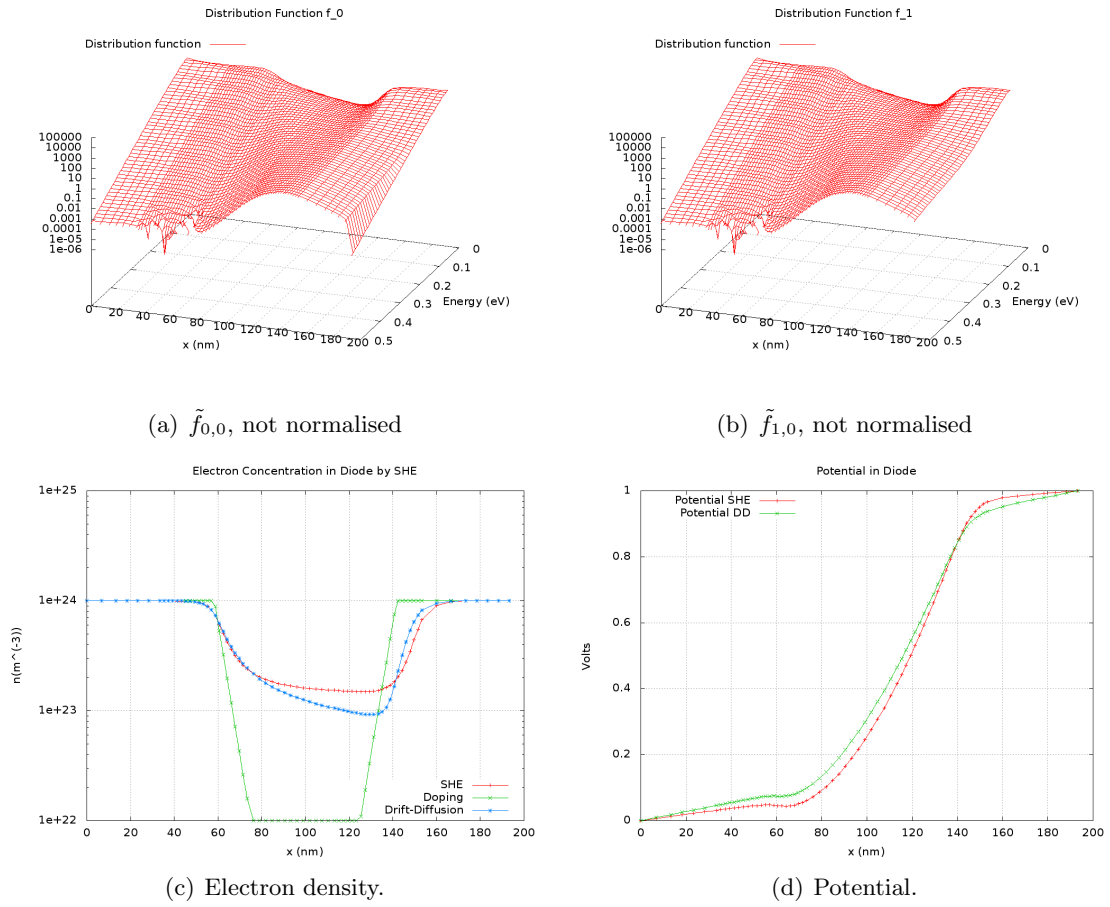


Figure 6.5: Self-consistent solution of SHE equations at an applied voltage of 1.0 V.

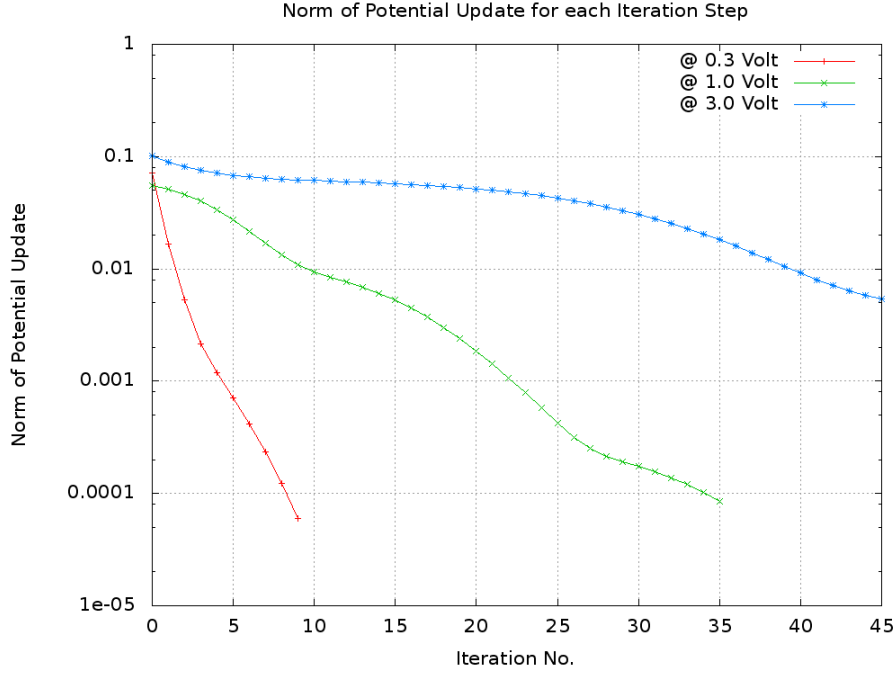


Figure 6.6: Norm of the potential update vector $\|\Delta\psi_s\|$ after each Gummel iteration step for different externally applied voltages.

The high number of Gummel iterations required for large externally applied voltages typically come with the need for a larger number of degrees of freedom due to the higher forces in the device. Thus, not only is the number of iterations higher, but also the time required for each iteration step. Faster convergence can be obtained by the use of the Newton method on the full coupled system as was observed recently [32]. In Fig. 6.7 the potential update $\Delta\psi_s$ according to the correction

$$\psi_{s+1} = \psi_s + d\Delta\psi_s \quad (6.27)$$

with damping factor d set to 0.8 is shown. One can see that another gain in convergence speed could be obtained by the use of multigrid methods for the nonlinear iteration: High frequency errors in the potential are eliminated through the Gummel iteration much faster than it is the case for low frequency errors. In particular, the potential updates are smooth functions, thus it is of advantage to start with a coarse mesh, perform several (fast) Gummel iterations to resolve low frequency updates of the potential and then interpolate the coarse numerical approximation to a finer mesh. Clearly, this requires stable discretisations in order to be able to obtain meaningful results on coarse meshes. For such a coarse mesh with only 30 grid points in each coordinate direction, the obtained simulation results are shown in Fig. 6.8. While the solution obtained from a 40×40 grid shows some numerical oscillations at higher energies only, the solution obtained from the 30×30 grid oscillates even at low energies and thus does not provide reliable numbers for the electron concentration.

To close this chapter, the importance of self consistent simulations, i.e. a full solution of the Boltzmann-Poisson system, is depicted in Fig. 6.9. The applied potential is 1 Volt, hence the associated potential can be found in Fig. 6.5. One can clearly see that small

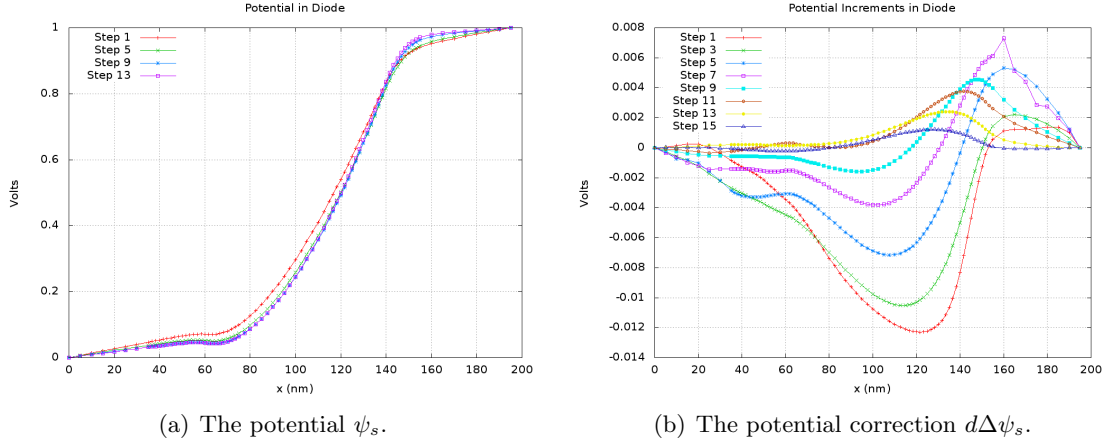


Figure 6.7: The potential and its correction at different Gummel iteration steps.

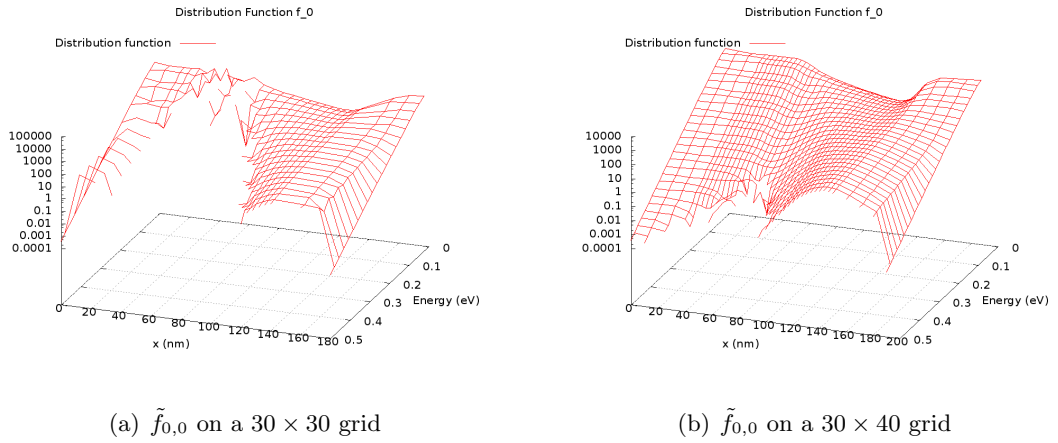


Figure 6.8: Coarse grid solutions for a n^+nn^+ diode at an applied voltage of 1.0 V.

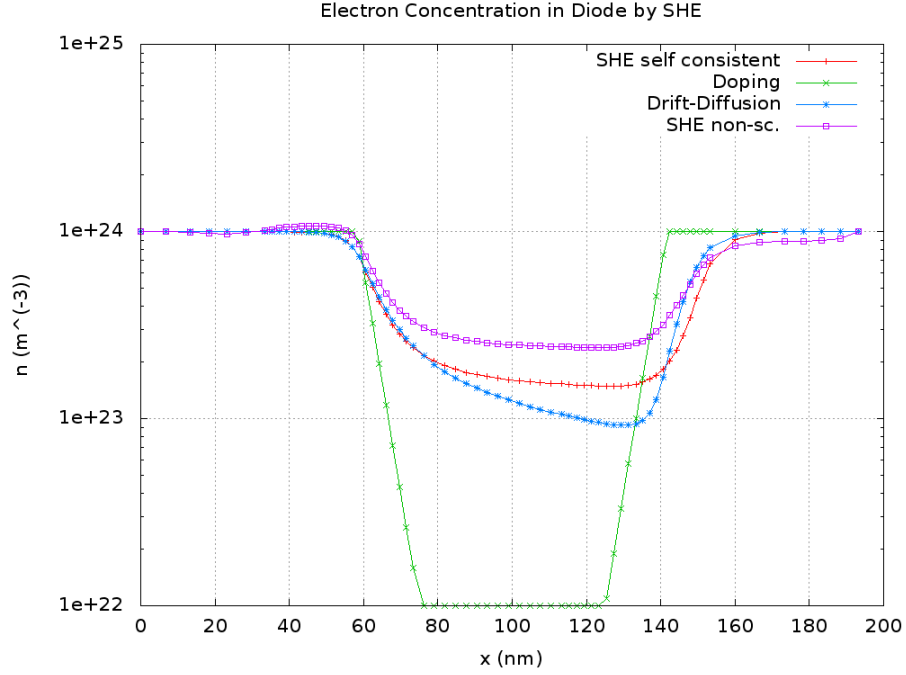


Figure 6.9: Comparison between electron concentrations obtained from self consistent, non self consistent and drift-diffusion simulations.

deviations of the potential in the order of a few percent lead to electron concentrations that differ by a factor of two in the case of a simple n^+nn^+ diode. For more complex devices such as bipolar transistors or MOSFETs, the difference is even higher. Consequently, one cannot rely on simpler (macroscopic) models for the computation of the potential and abandon the nonlinear iterations.

Chapter 7

Outlook and Conclusion

Several hints have already been dropped throughout this thesis on future investigations and possible extensions. Within this final chapter we sketch possible future work on the spherical harmonics expansion (SHE) of the Boltzmann transport equation (BTE) in a little more detail. The following sections give a brief overview of current ideas and hint at certain challenges which have to be met. However, one must not expect to find any fully developed theory or even results on the next pages. A conclusion is finally drawn in the closing section of this work.

7.1 Unstructured and Adaptive Grids

The simulation of the n^+nn^+ diode from Chapter 6 was carried out using a non-equidistant Cartesian tensor product grid with the different number of grid points in x and ε direction. The results show that there is only a minor deviation from the Maxwellian shape of the solution near the contacts, thus it would be sufficient to have a lower grid point density there. On the other hand, in the transition region between the highly doped n^+ region and the lightly doped n regions, a much higher density of grid points is desirable to obtain a good accuracy of the numerical solution.

A good indication of strong variations in the solution is given by the force term \mathbf{F} : In regions where the force is relatively small, electrons hardly respond to the force, whereas in areas of large forces electrons are accelerated and gain a considerable amount of energy. Such large forces typically show up at the border of regions with different doping levels, hence we expect to see a strong variation of the distribution function in \mathbf{x} -direction, while there the variation in energy is typically lower than for the Maxwellian shape at the contacts. The reason is that electrons do not gain energy instantly, they rather gain energy while they travel across the regions with high forces. This energy is then lost in the regions with lower forces due to scattering.

For a cell T with characteristic length h_T and centre at $(\mathbf{x}_T, \varepsilon_T)$ one may set

$$h_T \sim \frac{\varepsilon_0 + \varepsilon_T}{F_0 + |\mathbf{F}(\mathbf{x}_T)|} , \quad (7.1)$$

where ε_0 and F_0 are positive parameters that have to be chosen appropriately. With this choice of cell sizes, the cells are smallest near zero energy in regions where the force is largest, which coincides with the observations from the simulation results of an n^+nn^+ diode.

It is possible to choose cell sizes according to (7.1) on both structured and unstructured grids. This allows the construction of a sufficiently fine initial mesh, where each cell T can then be adaptively refined according to an error estimator $\eta(T, f_h)$ based on the initial numerical solution f_h . However, due to the typically exponential decay of f_h with energy, the use of simple standard error estimators based on the gradient of f_h are not expected to yield good results. Thus, using standard error estimators on f_h , the mesh is likely to be refined near zero energy only. Better results can be expected when standard error estimators are applied to $\ln(f_h)$, $f_h \exp((\varepsilon + q\psi)/(k_B T))$ or $\ln(f_h) + (\varepsilon + q\psi)/(k_B T)$. Such an adaptive mesh refinement leads to unstructured grids in a natural way, which comes at the cost of additional memory needed for mesh data structures. However, since for a SHE of order L , $(L + 1)^2$ unknowns are associated with each mesh node, the amount of memory needed for mesh handling is small compared to the memory needed for the unknowns as soon as L exceeds, say, five.

Another reduction of unknowns can be achieved by a clever selection of the energy boundary. We recall that macroscopic quantities such as electron concentration, average electron velocity or current density within the device are obtained from an integration over energy. For example, the macroscopic density u is computed by

$$u(\mathbf{x}, t) = \int_0^\infty \int_\Omega U(\varepsilon, \theta, \varphi) \tilde{f}(\mathbf{x}, \varepsilon, \theta, \varphi, t) d\Omega d\varepsilon. \quad (7.2)$$

In case $U \equiv Y_{0,0}$ one obtains the electron concentration, whereas the average energy can be obtained from the choice $U \equiv \varepsilon Y_{0,0}$. Since the integral along energy can only be evaluated numerically between zero and ε_{\max} due to the finite simulation domain, one can choose $\varepsilon_{\max} = \varepsilon_{\max}(\mathbf{x})$ in such a way that the resulting error is below a prescribed error. This is always possible, because the distribution function decays exponentially (at least for sufficiently large energies), while the function U is of polynomial growth for all microscopic quantities of interest. A potential pitfall is the specification of appropriate boundary conditions at $\varepsilon = \varepsilon_{\max}(\mathbf{x})$, because then the normal vectors of the boundary of the simulation domain do not point into energy direction only.

7.2 Adaptive Control of the Expansion Order

In Chapter 1 we have seen that the magnitude of the coefficients of the SHE decreases with increasing order. According to (1.94) the rate of decay is the larger the smoother the distribution function is. In terms of physics, a faster decay of the coefficients can be expected if only small voltages are applied to the device. This corresponds to results reported in the literature, where higher order expansions up to about degree nine are required to obtain good agreement with reference solutions computed using the Monte-Carlo method [24].

The decay of the coefficients of higher order harmonics can also be seen in results for a n^+nn^+ diode reported by Rahmat et. al. [33]. The crucial observation is that the higher order coefficients are largest in high-field regions, thus it may be sufficient to include these higher order coefficients only there and set them to zero in the low-field regions, where the distribution function is close to a Maxwellian. More explicitly, for

$$f(\mathbf{x}, \mathbf{k}, t) = \sum_{l=0}^L \sum_{m=-n}^n f_{n,m}(\mathbf{x}, \varepsilon) Y_{l,m}, \quad (7.3)$$

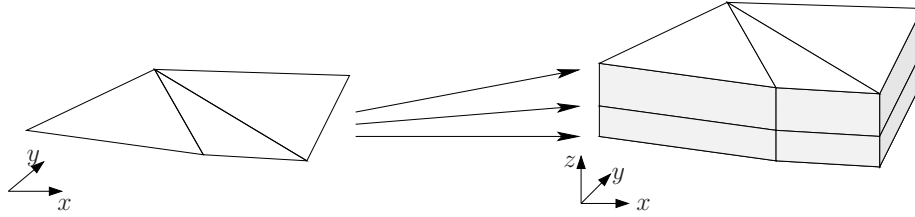


Figure 7.1: A three-dimensional prismatic mesh can be obtained from a two-dimensional mesh by appropriate translations. The two-dimensional cells are then facets of the three-dimensional prism. Here, the remaining facets are obtained by joining corresponding vertices.

and a given force \mathbf{F} , we set a-priori

$$f_{n,m}(\mathbf{x}, \varepsilon) \equiv 0, \quad \forall \mathbf{x} \in \Xi^{\text{low}} := \{\mathbf{y} \in \Xi : |\mathbf{F}(\mathbf{y})| < F_{l,m}^{\min}\}, \quad (7.4)$$

where Ξ denotes the simulation domain. It is to be expected that the threshold $F_{l,m}^{\min}$ has to be chosen for each l and m independently and that abrupt changes of the expansion order have to be avoided. At present it is not clear in what way the quality of the numerical solution will suffer from forcing coefficients to zero in the low-field region. Additionally, extra care has to be taken in order to preserve current continuity.

7.3 Higher Spatial Dimensions

The simulation of one-dimensional devices is certainly unsatisfactory considering the state of the art in device manufacture. A natural requirement thus is to simulate full three-dimensional devices, which actually requires the solution in a four-dimensional coordinate system, namely three spatial coordinates and one energy coordinate. In such a four-dimensional space, several obstacles pop up:

- *Meshing:* While a lot is known about meshing in two and three dimensions, hardly any concepts or software exist for meshing in four dimensions. Nevertheless it is still possible to obtain a full four-dimensional mesh by embedding a three-dimensional mesh of the device into the four-dimensional space. In particular, every mesh node located at \mathbf{x} in the three-dimensional mesh is embedded in the four-dimensional space at location $(\mathbf{x}, 0)$. A four-dimensional mesh of prismatic cells can then be obtained by repeatedly shifting the three-dimensional mesh along the energy axis to obtain the points $(\mathbf{x}, \Delta\varepsilon)$, $(\mathbf{x}, 2\Delta\varepsilon)$, $(\mathbf{x}, 3\Delta\varepsilon)$, \dots , $(\mathbf{x}, \varepsilon_{\max})$. This procedure is illustrated in Fig. 7.1 for the case of a transition from two to three dimensions.
- *Visualisation:* The visualisation of scalar quantities within a three-dimensional device often requires the use of iso-surfaces and other tools because of visibility problems. The visualisation of a four-dimensional energy distribution function is already a problem for the human perception, thus it will only be possible to visualise averaged quantities such as the particle concentration or the average energy of electrons at a certain point in the device. However, since the characteristic relationship of an electronic device is the current at various applied voltages, the inability to visualise the full four-dimensional energy distribution function should be a minor concern.

- *Computational Complexity:* The numerical example of a simplistic one-dimensional n^+nn^+ diode already showed that the system matrix carries some unpleasant properties such as a very high condition number. It is not clear whether the situation gets even worse in higher dimensions. Some authors already reported successful simulations of two-dimensional devices [16,20,23], but there have not been any publications on the simulation of full three-dimensional devices using SHE. Moreover, computational resources might (at present) be insufficient to use expansions of higher order. In a four-dimensional simulation domain, a uniform mesh refinement increases the number of unknowns by a factor of 16, therefore unstructured, adaptive grids as well as an adaptive control of the expansion order appear to be necessary to keep the number of unknowns to the absolute minimum.

7.4 Mathematical Modelling

The numerical simulation of an electronic device by SHE of different order has been the topic of this thesis. Early publications covered first order expansions only, which have then found their way into the world of mathematical modelling. There, the *mathematical SHE model* has been derived from scaling arguments and an Hilbert expansion of the distribution function [4]: The scattering operator Q is split into an elastic part Q_0 and a small perturbation Q_1 as

$$Q\{f^\alpha\} = Q_0\{f^\alpha\} + \alpha^2 Q_1\{f^\alpha\} , \quad (7.5)$$

where f^α denotes the scaled distribution function with respect to the substitutions $\mathbf{x}' = \alpha\mathbf{x}$ and $t' = \alpha t$. With this scaling, the BTE becomes

$$\frac{\partial f^\alpha}{\partial t} + \frac{1}{\alpha}(\mathbf{v} \cdot \nabla_{\mathbf{x}} f^\alpha + \mathbf{F} \nabla_{\mathbf{k}} f^\alpha) = \frac{1}{\alpha^2} Q_0\{f^\alpha\} + Q_1\{f^\alpha\} . \quad (7.6)$$

Substitution of a Hilbert expansion of the form

$$f^\alpha = f_0 + \alpha f_1 + \alpha^2 f_2 + \dots \quad (7.7)$$

and balancing terms of the same order in α leads to

$$Q_0\{f_0\} = 0 , \quad (7.8)$$

$$Q_0\{f_1\} = \mathbf{v} \cdot \nabla_{\mathbf{x}} f_0 + \mathbf{F} \nabla_{\mathbf{k}} f_0 , \quad (7.9)$$

$$Q_0\{f_2\} = \frac{\partial f^\alpha}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} f_1 + \mathbf{F} \nabla_{\mathbf{k}} f_1 - Q_1\{f_0\} . \quad (7.10)$$

Without going into the details, under suitable assumptions one can show that the above system is solvable if f_0 fulfils

$$\frac{\partial f_0}{\partial t} g(\varepsilon) + \nabla_{\mathbf{x}} \cdot \mathbf{j} + \mathbf{F} \cdot \frac{\partial \mathbf{j}}{\partial \varepsilon} = S\{f\} , \quad (7.11)$$

where

$$\mathbf{j} = \int_0^\infty \mathbf{v} f_1 g \, d\varepsilon = -\mathbf{D}(\mathbf{x}, \varepsilon) \left(\nabla_{\mathbf{x}} f_0 + \mathbf{F} \nabla_{\mathbf{k}} f_0 \right) , \quad (7.12)$$

$$S\{f\} = \int_0^\infty Q_1\{f\} g \, d\varepsilon , \quad (7.13)$$

with the density of states $g(\varepsilon)$ and the diffusion matrix $\mathbf{D}(\mathbf{x}, \varepsilon)$. A comparison of (7.11) with (3.28) shows that the latter carries an extra term $\mathbf{F} \cdot \mathbf{\Gamma}$. The reason for this difference can be traced back to the partial integration with respect to \mathbf{k} in (3.13): While the spherical harmonic $Y_{l,m}$ is a function of the angles, the scaling parameter α was assumed to be a scalar.

From the numerical point of view, (7.11) is easier to handle than (3.13), because after substitution of (7.12) into (7.11) one obtains a (parabolic) second-order partial differential equation, where the drift and diffusion terms can be clearly identified. We note that it is also possible to recast (6.14) into a second order PDE under some assumptions on the scattering operator: If we assume in (6.14) that $Q_0\{f\} = c_0 f_0$ and $Q_1\{f\} = c_1 f_1$, then elimination of f_1 leads to a PDE of second order for f_0 .

In contrast to the SHE model derived by scaling arguments, the set of equations (3.30) for the coefficients $f_{l,n}(\mathbf{x}, \varepsilon)$ still represent the full series expansion of the solution of the BTE. From the mathematical point of view it is interesting to ask in which sense the series expansion converges to the true solution f . For the diffusion matrix $\mathbf{D}(\mathbf{x}, \varepsilon)$ used in (7.12) the non-negativity has already been shown [4], thus one can check whether similar statements for the arbitrary-order equation (3.30) can be derived.

By different scaling arguments it is possible to derive the energy transport model, the hydrodynamic model and the drift-diffusion model from the BTE [4]. If we suppose that the set of equations (3.30) is only a recast BTE, then it might also be possible to derive macroscopic models from it. In particular, a closer investigation of the third-order model might lead to a new macroscopic model superior to for example the energy transport model.

7.5 Appropriate Linear Solver

Due to the additional energy coordinate, the discretisation of the SHE equations leads to a high number of unknowns. The resulting system matrix, however, is extremely sparse, hence iterative solvers should be chosen. Unfortunately, iterative solvers failed to converge for SHE in the low-scattering regime as reported in Chapter 6, thus a direct Gauss solver was used to obtain the results shown in the preceding chapter. However, direct solvers are inferior to iterative methods (provided that they converge) for sparse matrices with large numbers of unknowns, say, more than a million, thus it is absolutely mandatory to find a suitable iterative solver for SHE.

Tightly connected with the choice of an iterative solver is the choice of an appropriate preconditioner. Jungemann et. al. [24] reported the successful use of GMRES with a rather advanced incomplete LU-preconditioner, but also mentioned failure of convergence if the preconditioner was too rough.

An open question is the use of the block structure (5.33) for preconditioners. If the full system matrix $\mathbf{S} \in \mathbb{R}^{uv \times uv}$ had a (Kronecker) block structure in the form

$$\mathbf{S} = \mathbf{U} \otimes \mathbf{V} \quad (7.14)$$

with $\mathbf{U} \in \mathbb{R}^{u \times u}$, $\mathbf{V} \in \mathbb{R}^{v \times v}$, then a preconditioner could be obtained from approximate inverses $\tilde{\mathbf{U}}^{-1} \in \mathbb{R}^{u \times u}$ and $\tilde{\mathbf{V}}^{-1} \in \mathbb{R}^{v \times v}$ as

$$\tilde{\mathbf{S}}^{-1} = \tilde{\mathbf{U}}^{-1} \otimes \tilde{\mathbf{V}}^{-1} \quad (7.15)$$

because of

$$\tilde{\mathbf{S}}^{-1}\mathbf{S} = (\tilde{\mathbf{U}}^{-1} \otimes \tilde{\mathbf{V}}^{-1})(\mathbf{U} \otimes \mathbf{V}) = (\tilde{\mathbf{U}}^{-1}\mathbf{U}) \otimes (\tilde{\mathbf{V}}^{-1}\mathbf{V}) \approx \mathbf{I}_u \otimes \mathbf{I}_v \approx \mathbf{I}. \quad (7.16)$$

Unfortunately, \mathbf{S} is according to (5.42) only a linear combination of matrices that are Kronecker products:

$$\mathbf{S} = \sum_{i=1}^1 6\mathbf{U}_i \otimes \mathbf{V}_i, \quad (7.17)$$

thus it is insufficient to compute the approximate inverses $\tilde{\mathbf{U}}_i^{-1}$ and $\tilde{\mathbf{V}}_i^{-1}$. One possibility to design a preconditioner is to select $\mathbf{S}^* \in \mathbb{R}^{uv \times uv}$ with $\mathbf{S}^* = \mathbf{U}^* \otimes \mathbf{V}^*$, $\mathbf{U}^* \in \mathbb{R}^{u \times u}$, $\mathbf{V}^* \in \mathbb{R}^{v \times v}$ such that

$$\mathbf{S}^* = \operatorname{argmin} \|\mathbf{S} - \mathbf{S}'\| \text{ for } \mathbf{S}' \in \mathbb{R}^{u \times u} \otimes \mathbb{R}^{v \times v}. \quad (7.18)$$

Then, the approximate inverse of \mathbf{S}^* can be assembled from the approximate inverses of \mathbf{U}^* and \mathbf{V}^* . However, it is not clear whether on the one hand a suitable \mathbf{S}^* exists at all, i.e. whether the error $\|\mathbf{S} - \mathbf{S}'\|$ can be made sufficiently small, and on the other hand whether \mathbf{S}^* can be computed in an inexpensive way.

7.6 Finite Element Methods

In order to ensure current continuity, the Finite Volume Method was chosen for the discretisation. However, current continuity came at the price of dropping the maximum entropy dissipation scheme (MEDS) for the even order equations. Hence, if current continuity is not of utmost importance, MEDS can also be applied to even order equations and different discretisation methods such as the Finite Element Method (FEM) [6, 50] can be applied. Moreover, a previously created finite element programming framework of the author [38, 39] can immediately provide features like higher order discretisations, multigrid methods or adaptivity.

One has to keep in mind that FEM works typically well for second-order elliptic or parabolic partial differential equations, while hyperbolic problems cause several additional problems. Some have been addressed and solved over the past decades, but remedies have often been tailored to a particular set of equations. Since the SHE equations are a system of (degenerate) first order hyperbolic partial differential equations, spurious oscillations may occur if the mesh is too coarse.

7.7 Conclusion

The numerical solution of the BTE using SHE was considered and extensively documented within this thesis. Both the mathematical and physical background was presented and the projection onto spherical harmonics was carried out in a general form. A Galerkin scheme was used for the discretisation and analytical formulae for the resulting integrals have been derived. Additionally, it was shown that higher order expansions suffer from high memory requirements unless the system matrix is stored in a compressed form proposed in this thesis.

Both the assembly and the solution of the resulting linear system turned out to be challenging: While the former required the evaluation of complicated integrals, the correct handling of several unknowns per mesh point and preprocessing steps such as the solution of the drift-diffusion model, the latter requires an advanced preconditioner. Standard preconditioners shipped with general purpose solver libraries did not converge. As a remedy, a direct Gauss solver was used in one spatial dimension to obtain the simulation results for a n^+nn^+ -diode. Unlike the Monte Carlo method, a solution of the SHE equations provides a deterministic way to compute the electron distribution function in a device but comes at the price of higher algorithmic and implementation complexity.

Even though the SHE of the BTE is mathematically much harder to handle than for example elliptic partial differential equations of second order, it offers many links for advanced numerical techniques such as adaptive or automatic mesh refinement and an adaptive control of the expansion order. With such advanced numerical methods and the ever-increasing computational resources, the SHE method can complement or even replace other macroscopic models such as the energy transport model or the hydrodynamic model for device simulation in the future.

Bibliography

- [1] N. Ben Abdallah, P. Degond, P. Markowich, and C. Schmeiser. High Field Approximations of the Spherical Harmonics Expansion Model for Semiconductors. *Z. Angew. Math. Phys.*, 52(2):201–230, 2001.
- [2] S. Balay, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. Curfman McInnes, B. F. Smith, and H. Zhang. PETSc Web page, 2009. <http://www.mcs.anl.gov/petsc>.
- [3] W. W. Bell. *Special Functions for Scientists and Engineers*. D. Van Nostrand, London, 1968.
- [4] N. Ben Abdallah and P. Degond. On a Hierarchy of Macroscopic Models for Semiconductors. *Journal of Mathematical Physics*, 37:3306–3333, July 1996.
- [5] J. Bey. *Finite-Volumen- und Mehrgitter-Verfahren für elliptische Randwertprobleme*. B. G. Teubner, 1998.
- [6] D. Braess. *Finite Elements. Theory, Fast Solvers and Applications in Solid Mechanics*. Cambridge University Press, 1997.
- [7] K. F. Brennan. *The Physics of Semiconductors*. Cambridge University Press, 1999.
- [8] R. Brunetti. A Many-Band Silicon Model for Hot-Electron Transport at High Energies. *Solid State Electronics*, 32:1663–1667, December 1989.
- [9] Schmeiser C. and A. Zwirchmayr. Galerkin Methods for the Semiconductor Boltzmann Equation. In *Proc. of the Third International Congress on Industrial and Applied Mathematics*, 1995.
- [10] Schmeiser C. and A. Zwirchmayr. Elastic and Drift-Diffusion Limits of Electron-Phonon Interaction in Semiconductors. 8(1):37–53, 1998.
- [11] P. J. Davis. *Interpolation and Approximation*. Ginn and Co., Boston, 1963.
- [12] P. Degond. *Some Current Topics on Nonlinear Conservation Laws*. Studies in Advanced Mathematics. American Mathematical Society, 2000.
- [13] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere*. Numerical Mathematics and Scientific Computation. Clarendon Press, Oxford, 1998.
- [14] W. Gautschi. *Orthogonal Polynomials*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2004.

- [15] A. Gnudi, D. Ventura, and G. Baccarani. One-dimensional Simulation of a Bipolar Transistor by means of Spherical Harmonics Expansion of the Boltzmann Transport Equation. In W. Fichtner and D. Aemmer, editors, *Proc. of SISDEP*, volume 4, pages 205–213, 1991.
- [16] A. Gnudi, D. Ventura, G. Baccarani, and F. Odeh. Two-dimensional MOSFET Simulation by Means of a Multidimensional Spherical Harmonics Expansion of the Boltzmann Transport Equation. *Solid-State Electronics*, 36(4):575–581, 1993.
- [17] N. Goldsman, L. Hendrickson, and J. Frey. A Physics-based Analytical/Numerical Solution to the Boltzmann Transport Equation for the Use in Device Simulation. *Solid State Electronics*, 34:389–396, 1991.
- [18] F. Golse and F. Poupaud. Limite fluide des équations de Boltzmann des semiconducteurs pour une Statistique de Fermi-Dirac. *Asympt. Anal.*, 6:135–160, July 1992.
- [19] H. Groemer. *Geometric Applications of Fourier Series and Spherical Harmonics*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1996.
- [20] K. A. Hennacy, N. Goldsman, and I. D. Mayergoyz. 2-Dimensional Solution to the Boltzmann Transport Equation to Arbitrarily High-Order Accuracy. In *Proceedings of IWCE*, pages 118–122, 1993.
- [21] K. A. Hennacy, Y. J. Wu, N. Goldsman, and I. D. Mayergoyz. Deterministic MOSFET Simulation using a Generalized Spherical Harmonic Expansion of the Boltzmann Equation. *Solid-State Electronics*, 38(8):1485–1495, 1995.
- [22] E. W. Hobson. *The Theory of Spherical and Ellipsoidal Harmonics*. Cambridge University Press, 1955.
- [23] S. M. Hong, C. Jungemann, and M. Bollhofer. A Deterministic Boltzmann Equation Solver for Two-Dimensional Semiconductor Devices. In *Simulation of Semiconductor Processes and Devices, 2008. SISPAD 2008. International Conference on*, pages 293–296, Sept. 2008.
- [24] C. Jungemann, A. T. Pham, B. Meinerzhagen, C. Ringhofer, and M. Bollhöfer. Stable Discretization of the Boltzmann Equation based on Spherical Harmonics, Box Integration, and a Maximum Entropy Sissipation Principle. *Journal of Applied Physics*, 100(2):024502–+, July 2006.
- [25] A. Jüngel. *Transport Equations for Semiconductors*. Lecture Notes in Physics No. 773. Springer, Berlin, 2009.
- [26] D. Levermore. Moment Closure Hierarchies for Kinetic Theories. *Journal of Statistical Physics*, 83:1021–1065, 1996.
- [27] W. Liang, N. Goldsman, and I. Mayergoyz. 2-D MOSFET Modeling Including Surface Effects and Impact Ionization by Self-Consistent Solution of the Boltzmann, Poisson and Hole-Continuity Equations. *IEEE Transactions on Electron Devices*, 44(2):257–267, 1997.

- [28] H. Lin, N. Goldsman, and I. D. Mayergoyz. Deterministic BJT Modeling by Self-Consistent Solution to the Boltzmann, Poisson and Hole-Continuity Equations. In *Proceedings of IWCE*, pages 55–59, 1993.
- [29] M. Lundstrom. *Fundamentals of Carrier Transport*. Cambridge University Press, second edition, 2000.
- [30] P. A. Markowich, C. A. Ringhofer, and C. Schmeiser. *Semiconductor Equations*. Springer, Wien, New York, 1990.
- [31] A. Okabe, B. Boots, and K. Sugihara. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, Inc., New York, NY, USA, 1992.
- [32] A. T. Pham, C. Jungemann, and B. Meinerzhagen. A Convergence Enhancement Method for Deterministic Multisubband Device Simulations of Double Gate PMOS-FET. In *Proceedings of SISPAD*, pages 115–118, 2009.
- [33] K. Rahmat, J. White, and D. A. Antoniadis. Simulation of Semiconductor Devices Using a Galerkin/Spherical Harmonic Expansion Approach to Solving the Coupled Poisson-Boltzmann System. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 15(10):1181–1195, 1996.
- [34] C. Ringhofer. Space-Time Discretization of Series Expansion Methods for the Boltzmann Transport Equation. *SIAM J. Numer. Anal.*, 38(2):442–465, 2000.
- [35] C. Ringhofer. Dissipative Discretization Methods for Approximations to the Boltzmann Equation. *Math. Mod. Meth. in Appl. Sci.*, 11:133–149, 2001.
- [36] C. Ringhofer. Numerical Methods for the Semiconductor Boltzmann Equation Based on Spherical Harmonics Expansions and Entropy Discretizations. *Transport Theory and Statistical Physics*, 31:431–452, January 2002.
- [37] C. Ringhofer. A Mixed Spectral-Difference Method for the Steady State Boltzmann-Poisson System. *SIAM J. Numer. Anal.*, 41(1):64–89, 2003.
- [38] K. Rupp. Solving PDEs in Electromigration using a Generic Template-Metaprogramming Framework. Master’s thesis, Brunel University, 2007.
- [39] K. Rupp. Multiphysics Modelling in the Context of Generative Programming . Master’s thesis, Institute for Microelectronics, TU Wien, 2009.
- [40] Y. Saad. *Iterative Methods for Sparse Linear Systems, Second Edition*. Society for Industrial and Applied Mathematics, April 2003.
- [41] D. L. Scharfetter and H. K. Gummel. Large-Signal Analysis of a Silicon Read Diode Oscillator. *IEEE Transactions on Electron Devices*, 16:64–77, 1969.
- [42] S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Springer, Wien, 1984.

- [43] S. P. Sing. *Modeling Multi-Band Effects of Hot-Electron Transport in Simulation of Small Silicon Devices by a Deterministic Solution of the Boltzmann Transport Equation Using Spherical Harmonic Expansion*. PhD thesis, Department of Electrical Engineering, University of Maryland, College Park, MD 20742, 1998.
- [44] W. J. Sternberg and T. L. Smith. *The Theory of Potential and Ellipsoidal Harmonics*. University of Toronto Press, 1952.
- [45] M. C. Vecchi, J. Mohring, and M. Rudan. An efficient Solution Scheme for the Spherical-Harmonics Expansion of the Boltzmann Transport Equation [MOS Transistors]. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 16(4):353–361, Apr 1997.
- [46] M. C. Vecchi and M. Rudan. Modeling electron and hole transport with full-band structure effects by means of the Spherical-Harmonics Expansion of the BTE. *IEEE Transactions on Electron Devices*, 45:230–238, January 1998.
- [47] M. C. Vecchi, D. Ventura, A. Gnudi, and G. Baccarani. Incorporating Full Band-Structure Effects in the Spherical Harmonics Expansion of the Boltzmann Transport Equation. In *Numerical Modeling of Processes and Devices for Integrated Circuits, 1994. NUPAD V., International Workshop on*, pages 55–58, Jun 1994.
- [48] D. Ventura, A. Gnudi, and G. Baccarani. A Deterministic Approach to the Solution of the BTE in Semiconductors. *Nuovo Cimento Rivista Serie*, 18:1–33, June 1995.
- [49] D. Ventura, A. Gnudi, G. Baccarani, and F. Odeh. Multidimensional Spherical Harmonics Expansion of Boltzmann Equation for Transport in Semiconductors. *Appl. Math. Lett.*, 5(3):85–90, 1992.
- [50] O. C. Zienkiewicz and R. L. Taylor. *The Finite Element Method - Volume 1: The Basis*. Butterworth-Heinemann, 5th edition, 2000.

Index

- associated Legendre function
 - first kind, 8
 - second kind, 8
- average energy, 35
- barycenter method, 51
- Bernoulli function, 54, 60
- Boltzmann transport equation, 25
- Boltzmann-Poisson system, 26
- Brillouin zone, 24
- carrier concentration, 33
- carrier density, 33
- circumcenter method, 50
- condensed SHE, 18
- Condon-Shortley phase, 8
- continuity equation, 55
- current density
 - generalised, 38
- damping factor, 99
- Delaunay triangulation, 50
- density of states
 - generalised, 33
 - physical, 34
- dispersion relation, 31
- distribution function
 - generalised energy, 37
- drift-diffusion model, 89, 96
- entropy, 44
- entropy functional, 44
- finite volume method, 49
- free streaming operator, 41
- Galerkin method
 - entropy based, 45
- GMRES, 92
- gradient
 - in spherical coordinates, 38
- Gummel scheme, 90
- high energy tail, 97
- Kronecker product, 83
- Krylov subspace method, 92
- Laplace equation, 3
- Legendre function
 - of the second kind, 6
- Legendre polynomial, 6
- Legendre's equation, 5
- Liouville's equation, 23
 - semiclassical, 24
- low density approximation, 29
- major index, 8
- Maxwellian, 91
- mean electron velocity, 34
- minor index, 8
- naive method, 66
- non-parabolicity factor, 31
- parabolic band approximation, 30
- Pauli principle, 26
- Poisson equation, 3
- preconditioner
 - diagonal, 93
 - exponential, 93
 - row scaling, 93
- rectangular grid, 58
- relaxation time approximation, 29
- scattering
 - carrier-carrier, 28
 - ionised impurity, 28
 - phonon, 28
- scattering operator
 - low density approximation, 39

scattering rate, 26
scattering term
 in-, 27, 39
 out-, 27, 39
Scharfetter-Gummel scheme, 52
self consistent solution, 26
sparsity indicator, 67
spherical Fourier transform, 18
spherical harmonics, 3, 12
strong formulation, 49

thermal voltage, 90

velocity randomising, 40
Vlasov equation, 25
Vlasov-Poisson system, 25
Voronoi diagram, 50