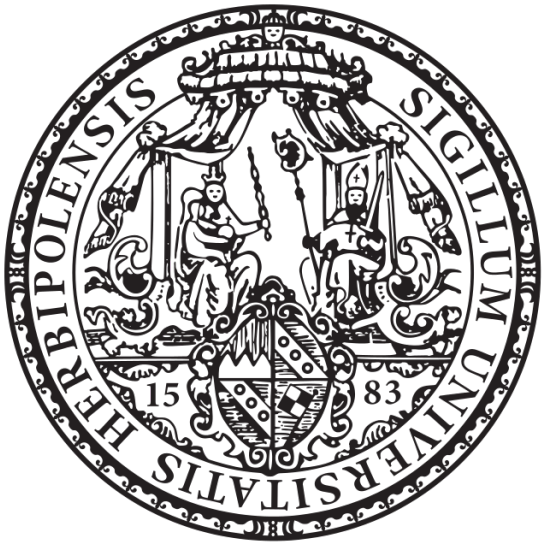
Spin Transport in Topological Insulators and Geometrical Spin Control

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Julius-Maximilians-Universität Würzburg



vorgelegt von

Dietrich Gernot Rothe

aus Tübingen

Würzburg 2015

Eingereicht am: 7.5.2015
bei der Fakultät für Physik und Astronomie
1. Gutachterin: Prof. Dr. Ewelina M. Hankiewicz
2. Gutachter: Prof. Dr. Giorgio Sangiovanni
der Dissertation.

Vorsitzender: Prof. Dr. Karl Brunner
1. Prüferin: Prof. Dr. Ewelina M. Hankiewicz
2. Prüfer: Prof. Dr. Giorgio Sangiovanni
3. Prüfer: Prof. Dr. Hartmut Buhmann
im Promotionskolloquium.

Tag des Promotionskolloquiums: 15.1.2016

Doktorurkunde ausgehändigt am: .....

Zusammenfassung

Manipulation und Transport von elektronischen Spins sind die wesentlichen Elemente, die für das Funktionieren einer zukünftigen Spin-basierten Elektronik implementiert werden müssen. Diese Arbeit befasst sich schwerpunktmäßig mit Halbleitersystemen, in denen diese Prinzipien mit hoher Zuverlässigkeit möglich sind. Dazu wurden sowohl numerische als auch analytische Berechnungsmethoden genutzt, letztere oft in der Form einfacher Modelle zur Interpretation der numerischen Ergebnisse.

Das Halbleitersystem von HgTe/CdTe Quantentrögen, auch bekannt als zweidimensionaler topologischer Isolator, ist sowohl von fundamentalem wissenschaftlichen Interesse, da die topologisch nichttriviale Energiestruktur zu einem Schutz von Transporteigenschaften führt, als auch von angewandterem Interesse, da aus diesem Materialsystem Proben gefertigt werden können, die ballistischen Transport hoher Qualität zeigen, und da zudem die Rashba Spin-Bahn-Kopplung sowie die elektronische Dichte durch elektrische Steuerelektroden einstellbar sind. Wir erweitern das Bernevig-Hughes-Zhang Modell für zweidimensionale topologische Isolatoren, indem wir ein Vierbandmodell herleiten, das Rashba Spin-Bahn-Kopplungsterme enthält, die durch ein äußeres elektrisches Feld hervorgerufen werden, wenn dieses die Inversionssymmetrie des Quantentroges bricht. Der Transport von Spins in diesem System zeigt ein interessantes Wechselspiel zwischen Effekten der Rashba Spin-Bahn-Kopplung und Effekten der intrinsischen Dirac-artigen Spin-Bahn-Kopplung. Dabei dominiert die Rashba Spin-Bahn-Kopplung das Verhalten des Spin-Hall-Signals. Basierend auf der einstellbaren Rashba Spin-Bahn-Kopplung, schlagen wir einen spinselektiven Polarisator zur rein elektrischen Erzeugung und Detektion von Spinströmen vor. Das Funktionsprinzip ist vergleichbar mit demjenigen eines doppelbrechenden Kristalls. In der vorgeschlagenen Anordnung untersuchen wir die Spinpolarisation in verschieden Spinvektorkomponenten und zeigen die Realisierbarkeit von hoher Spinpolarisation in der Ebene. Da der Spin keine Erhaltungsgröße des Halbleitermodells ist, analysieren wir in einem ersten Schritt den Transport von der Erhaltungsgröße Helizität, und setzen die erzeugte Polarisation dann in Bezug zur Spinpolarisation.

Des Weiteren analysieren wir thermoelektrischen Transport in einem System, das auch den Spin-Hall-Effekt zeigt. Aufgrund von Spin-Bahn-Kopplung kommt es beim Anlegen eines Temperaturgradienten zu einem transversalen Spinstrom, genannt Spin-Nernst-Effekt. Dieser ist über eine Mott-artige Beziehung mit dem Spin-Hall-Effekt verknüpft. Im metallischen Energiebereich können wir die Signale qualitativ anhand von einfachen analytischen Modellen verstehen. Im Energiebereich der elektronischen Bandlücke finden wir ein Spin-Nernst-Signal, das vom räumlichen Überlapp der Randzustände herrührt, die an gegenüberliegenden Kanten des Halbleitersystems lokalisiert sind.

Im methodischen ersten Teil dieser Arbeit diskutieren wir zwei komplementäre Methoden zur Konstruktion von effektiven Halbleitermodellen, nämlich die Methode der Envelopefunktionen und die Methode der Invarianten. Außerdem präsentieren wir Elemente der elektronischen Transporttheorie, unter besonderer Beachtung von Spintransport. Wir diskutieren die Zusammenhänge zwischen dem adiabatischen Theorem in der Quantenmechanik einerseits, und semiklassischer Transporttheorie sowie der topologischen Klassifizierung von Phasen andererseits. Als weitere Anwendung des adiabatischen Theorems zeigen wir, wie universelle Kontrolle eines einzelnen Spins in einem Quantenpunkt aus Schwerlochzuständen experimentell realisiert werden kann, ohne dabei die Zeitumkehrsymmetrie zu brechen. Zu diesem Zweck führen wir ein elektrisches Quadrupolfeld ein, dessen Konfiguration als adiabatischer Kontrollparameter dient. Wir schlagen die experimentelle Realisierung des Quantenpunktes in einem QaAs/GaAlAs Quantentrogsystem vor.

Summary

In the field of spintronics, spin manipulation and spin transport are the main principles that need to be implemented. The main focus of this thesis is to analyse semiconductor systems where high fidelity in these principles can be achieved. To this end, we use numerical methods for precise results, supplemented by simpler analytical models for interpretation.

The material system of 2D topological insulators, HgTe/CdTe quantum wells, is interesting not only because it provides a topologically distinct phase of matter, physically manifested in its protected transport properties, but also since within this system, ballistic transport of high quality can be realized, with Rashba spin-orbit coupling and electron densities that are tunable by electrical gating. Extending the Bernvevig-Hughes-Zhang model for 2D topological insulators, we derive an effective four-band model including Rashba spin-orbit terms due to an applied potential that breaks the spatial inversion symmetry of the quantum well. Spin transport in this system shows interesting physics because the effects of Rashba spin-orbit terms and the intrinsic Dirac-like spin-orbit terms compete. We show that the resulting spin Hall signal can be dominated by the effect of Rashba spin-orbit coupling. Based on spin splitting due to the latter, we propose a beam splitter setup for all-electrical generation and detection of spin currents. Its working principle is similar to optical birefringence. In this setup, we analyse spin current and spin polarization signals of different spin vector components and show that large in-plane spin polarization of the current can be obtained. Since spin is not a conserved quantity of the model, we first analyse the transport of helicity, a conserved quantity even in presence of Rashba spin-orbit terms. The polarization defined in terms of helicity is related to in-plane polarization of the physical spin.

Further, we analyse thermoelectric transport in a setup showing the spin Hall effect. Due to spin-orbit coupling, an applied temperature gradient generates a transverse spin current, i.e. a spin Nernst effect, which is related to the spin Hall effect by a Mott-like relation. In the metallic energy regimes, the signals are qualitatively explained by simple analytic models. In the insulating regime, we observe a spin Nernst signal that originates from the finite-size induced overlap of edge states.

In the part on methods, we discuss two complementary methods for construction of effective semiconductor models, the envelope function theory and the method of invariants. Further, we present elements of transport theory, with some emphasis on spin-dependent signals. We show the connections of the adiabatic theorem of quantum mechanics to the semiclassical theory of electronic transport and to the characterization of topological phases. Further, as application of the adiabatic theorem to a control problem, we show that universal control of a single spin in a heavy-hole quantum dot is experimentally realizable without breaking time reversal invariance, but using a quadrupole field which is adiabatically changed as control knob. For experimental realization, we propose a GaAs/GaAlAs quantum well system.

Contents

Chapter 1

Introduction

Mesoscopic physics can be defined as the study of physical systems of sub-micron or nanoscale, which are small enough to reveal quantum mechanical interference effects, yet large enough to allow the use of statistical concepts like chemical potential and temperature. Therefore, mesoscopic systems are the ideal candidates for studying the transition of classical to quantum mechanical behaviour. Since microstructured semiconductors provide an ideal way to implement mesoscopic devices, mesoscopic physics can be regarded as subfield of semiconductor and solid state physics. In quantitative terms, the phase coherence length l_ϕ of a mesoscopic device should be comparable to the extension of the device [Dat07; Bee+91]. Usually, scattering events that cause decoherence will also change the momentum, so the mean free path l_0 will be smaller than l_ϕ. Transport in very clean samples, where barely any scattering at disorder happens inside the sample, is called ballistic, and these devices will also show quantum interference effects. In this context, the most prominent experimental achievement has been to build a microsctructure showing the Aharonov-Bohm effect, i.e. interference due to enclosed magnetic flux in a ring [Web+85]. Certainly, one experimental direction of scientific effort is to find larger and larger systems which still show quantum mechanical behaviour.

Interestingly, the simplest phenomenological physical laws known from school books, like Ohm's law for electrical conductance and Faraday's law for heat conductance, no longer hold on the mesoscopic scale. For this reason, it has been a tough problem to find microscopic derivations of these laws, based only on quantum electrodynamics and statistical physics, and a key point in these derivations is to assume sample dimensions large enough to allow for quantum mechanical phase decoherence due to interactions, at a scale small compared to the sample's extension [Dub+11; Dat07]. Note that we can already say by the second law of thermodynamics, that Ohm's law cannot hold in a mesoscopic system without decoherence. Thermodynamics relates dissipation, which is conversion of work to heat, to a loss of information about the microscopic state. No decoherence means no loss of information about the system, so dissipation cannot take place, and by contrast a finite DC conductivity as in Ohm's law implies dissipation[1]. Some care must be taken with this simple argument, since we have not taken into account that transport usually requires an open systems (i.e. with contacts to the environment). Dissipation at the contacts will still be present for the phase-coherent mesoscopic system. Another difficulty in such theoretical efforts is that chemical potential and temperature are strictly defined only in equilibrium. However, transport of heat or charge is a non-equilibrium problem. Since local

[1] An exception can occur when driving electrical field and induced current are orthogonal, like in the quantum Hall state.

chemical potential and local temperature are very intuitive and useful concepts which appear in the mentioned phenomenological laws for transport, there have been attempts to find strict definitions for these. The problem of heat transport at the microscopic scale turns out to be even more difficult than the problem of charge transport. There are mainly two reasons for this. Firstly, there are more physical effects that contribute to heat transport - phonons and electrons, neglecting radiation. Secondly, while the charge is a conserved quantity, heat is not - from thermodynamics we know that the conserved quantity, energy, splits up in the parts heat and work. This has to be taken into account for a microscopically valid definition of heat flow.

So far, we have only mentioned charge and heat currents, but this thesis is more concerned with yet another kind of current, connected with the transport of electron or quasi-particle spins. Interest in spin transport has been raised by the idea that for information processing, neither heat nor charge transport is needed - instead, one needs just some property that can be manipulated and transported. Since the electron intrinsically has a spin degree of freedom, this insight has spawned the field of spintronics, which deals with spin-based information processing. Therefore, generation and detection of spin currents is one of the major goals of spintronics [Wol+01].

Considering the advanced level of current information processing technology, one should have strong arguments if one proposes to shift away from established charge-based information processing to something different. There are some fundamental advantages of spin-based information processing. The technological advancement in information processing of the last decades has been primarily due to scaling down of structures, in order to fit more logical gates on a semiconductor system. This naturally comes with increased heat dissipation, and nowadays, this has become the main impediment for further advances in computation power. If microstructures are further scaled down, quantum mechanical effects will naturally appear. Instead of viewing this only as difficulty, we want to point out that quantum mechanics can also be used as working principle for new logical devices.

Spin-based information processing can have significant benefits in terms of energy per switching process [Aws+07]. Firstly, the fundamental thermodynamic limit of $k_B T \log 2 \approx 25$ meV per charge based information processing of one bit does not apply to spin-based technology, yet this limit is far from being reached in current technology. More importantly, switching speed of charge based bit registers is limited by capacitance. Moore's empirical law of exponential technological advancement in the number of logical gates per device still seems to hold today. However, typical processor clock speeds no longer increase from generation to generation these days, and have stayed in the GHz regime over the last decade. By contrast, the switching speed limitation in spintronic devices is given by the spin precession frequency, which can be of the order of THz. Combining spin-based information processing with conventional electronics will extend technological possibilities, in particular for optical and magnetic working principles. The hope is to use the magnetic moment of the spin for integration of information processing and persistent information storage. In current technology, the latter is still predominately done using magnetic principles. Further, spin qubits are promising candidates for implementation of quantum computation [Los+98], due to achievable long coherence times. Spin qubits embedded in semiconductors can provide scalability and the possibility of integration with conventional electronics on a level that is unmet by other qubit implementations.

Another field which has attracted a lot of interest in mesoscopic physics, is the discovery of topologically non-trivial phases of matter. Classifying phases of matter has always been one of the main interests of physicists. Until the discovery of the quantum Hall effect in 1980 [Kli+80],

it had been thought that all possible thermodynamic phases could be classified on the basis of broken symmetries. Instead, the quantum Hall state is classified by an invariant that depends on the topology of electronic wave functions. A strong magnetic field is required for realization of the quantum Hall effect, so time reversal symmetry is broken in this state. Recently, a topologically non-trivial phase has been discovered where time reversal symmetry is not broken, known as the quantum spin Hall state [Kan+05a; Ber+06a; Kön+07]. Instead of strong magnetic fields, strong spin-orbit coupling is a necessity for the observation of the quantum spin Hall effect. Both the quantum Hall and the quantum spin Hall effect show localized edge states which contribute to dissipationless charge transport, and which appear as fingerprint of the non-trivial bulk state topology.

For technological applications, robustness against perturbations like disorder caused by lattice imperfections is very important, and operation at room temperature will usually be the goal. This makes topological protection of transport a very interesting feature. In this context we would also like to mention the recent discovery of graphene [Nov+05b]. Graphene, a two-dimensional layer of carbon consisting of a single atomic layer, shows features of relativistic physics as well as interesting transport properties like high electron mobility, and even the quantum Hall effect at room temperature [Nov+07]. Since the topological protection of charge transport in quantum spin Hall systems does not require magnetic fields, it could be even more interesting for applications, than the quantum Hall state. However, the quality of the edge states' protection is not comparable to the quality of the chiral quantum Hall edge states, since missing spatial separation of counter-propagating states means that there is no protection against higher-order processes like inelastic backscattering or two-particle backscattering.

So far, we have discussed fundamental scientific and technological motivations for research in mesoscopic physics. A third motivation for this kind of research can be interest in the methods for solving difficult physical problems. Methods are the main matters of part I of this thesis. Building effective models can be much more than the "cooking down" of microscopically strictly valid models, since often, physical behaviour at larger scale, by which we mean a larger number of particles, can be very different and much richer than what is seen as behaviour of individual, however interacting particles. Like Anderson put it, "more is different" [And72]. New phases of matter are the perfect example for this, since strict classification by broken symmetry or by topology can only work in the limit of infinitely large systems. I believe that research in the field of mesoscopic physics especially benefits from the application and extension of different effective models, in order to view physical behaviour from different angles. Effective models may be suited more for analytical or numerical analysis, but there is no strict line.

We will explain different approaches to find effective models, by controlled approximation or by working with a set of symmetries. This includes the envelope function method, which is a generalization of $\mathbf{k} \cdot \mathbf{p}$ theory for crystals, and which is outlined in Chapter 2, the method of invariants, which is presented in Chapter 3, and the adiabatic theorem of quantum mechanics and some applications thereof, which are treated in Chapter 4. The basis of electronic theory transport is explained in Chapter 5, and it will be used for the numerical implementation of transport calculations. Some details connected with the application of transport theory to spin-dependent problems are given in Appendices C, D, E and F. Chapter 6 gives a short introduction to topologically non-trivial semiconductor systems. The geometrical features of adiabatic transport are closely connected to a system's topology.

In part II of this thesis, we continue with applications of the set of methods of the theory part to experimentally realizable systems. In Chapter 7, we derive an extension of the Bernevig-

Hughes-Zhang (BHZ) model [Ber+06a] for the 2D topological insulator system of HgTe quantum wells. We include an inversion breaking potential which gives rise to Rashba-type spin splitting of the quantum well band structure, and we analyse numerically the interplay of different spin-orbit terms in spin transport. The Rashba spin splitting in the system of HgTe quantum wells can be large, and it is particularly interesting because it is electrically tunable. Further, gating of the system allows for different transport regimes to be analysed.

Chapter 8 deals with the control of a single spin of a heavy hole trapped in a quantum dot, by application of adiabatic transport. Therefore and in contrast to the other chapters, this chapter is relevant to proposals for building a quantum computer. We propose how to experimentally realize an all-electrical universal single-spin gate, which could be used in a quantum computer.

Interestingly, the same class of materials that reveal topological features often are good materials for thermoelectrics applications. In Chapter 9 we are interested in the relation of spin currents and heat currents in HgTe quantum wells. Different regimes of the material of 2D topological insulator are considered, and finite size effects are shown to appear in the relation of spin current and heat current.

Generation of spin currents is still one main technological difficulty that needs to be overcome for the future of spintronics. Using ferromagnetic constituents for this purpose seems to be connected with a lot of technological difficulties. Since it is desirable to find realizations of spin current injection based on all-electrical principles, we propose an all-electrical beam splitter setup for generation and detection of a spin current in Chapter 10. This is similar to an idea by Khodas [Kho+04], but we analyse a device of realistic geometry and analyse spin current and spin polarization signals of different spin directions. This chapter builds on the extended BHZ model of Chapter 7, which introduces competing spin-orbit terms.

In Chapter 11 we close with conclusions and an outlook on interesting open questions.

Part I

Theory

Chapter 2

Envelope function theory

The $\mathbf{k}\cdot\mathbf{p}$ method for obtaining multiband Hamiltonians for crystals is old and well established [Bas88; Win05]. The envelope function method is a natural generalization to problems where strict lattice-periodicity of a potential is no longer present, like in semiconductor heterostructures. Since the envelope function method is an approximation, there have been different approaches to the development of the theory, which may be incompatible in detail [Bas88; Bur88b; Bur99]. The most useful approach seems to be that of Burt [Bur88b]. The importance of the method is due to its high flexibility while at the same time, this method is simple enough to allow for intuitive physical interpretation.

2.1 $\mathbf{k}\cdot\mathbf{p}$ theory

Within $\mathbf{k}\cdot\mathbf{p}$ theory, the band structure of a crystal can be calculated to arbitrary precision, in principle. For practical purposes, one is merely interested in a Hamiltonian of small matrix dimension giving a good local approximation of the band structure. The properties of electronic transport are determined by states near the Fermi wave vector k_F. Usually, k_F is in the vicinity of a point of high symmetry, since in the intrinsic bulk semiconductor, the conduction band is empty. So instead of calculating the band structure in the full Brillouin zone, an approximation scheme to determine the band structure near a point $\mathbf{k}_0$ of high symmetry is sufficient. We will consider only the Γ point, i.e. $\mathbf{k}_0 = 0$, but the $\mathbf{k}\cdot\mathbf{p}$ method also works for $\mathbf{k}_0 \neq 0$, rendering the equations a bit more complicated.

For a bulk crystal, the Schrödinger equation including spin-orbit (SO) coupling reads

$$\left(\frac{p^2}{2m_0} + V(\mathbf{r}) + \frac{\hbar}{4m_0^2c^2}\left(\boldsymbol{\sigma}\times\nabla V(\mathbf{r})\right)\cdot\mathbf{p}\right)\psi(\mathbf{r}) = E\,\psi(\mathbf{r}) \tag{2.1}$$

where $V(\mathbf{r})$ is the crystalline potential with Bravais lattice periodicity, m_0 is the bare electron mass and c is the speed of light. $\boldsymbol{\sigma}$ is the vector of Pauli spin matrices. According to the Bloch theorem, a solution $\psi(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}}u_{n,\mathbf{k}}(\mathbf{r})$ can be written in terms of a lattice-periodic function $u_{n,\mathbf{k}}(\mathbf{r})$ and a plane wave prefactor where the crystal momentum $\mathbf{k}$ is in the first Brillouin zone. By using the Bloch wave function and evaluating the operator $\mathbf{p} = -i\hbar\nabla$ in (2.1), we obtain an equation for the part $u_{n,\mathbf{k}}(\mathbf{r})$. For fixed $\mathbf{k}_0$, the set of $\langle\mathbf{r}|n,\mathbf{k}_0\rangle = u_{n,\mathbf{k}_0}(\mathbf{r})$ provides a complete basis of lattice-periodic functions. Here n is a combined band and spin index. Thus we may expand $|n,\mathbf{k}\rangle = \sum_m c_m(\mathbf{k})|n,0\rangle$. At this point we make the non-essential assumption that the $\{|n,0\rangle\}_n$ are a basis without SO - which can conveniently be included later as a perturbation.

We obtain the matrix eigenvalue equation for the expansion coefficients where $\mathbf{k}$ enters only parametrically,

$$\sum_m \left(\delta_{lm}(E_m(0) + \frac{\hbar^2 k^2}{2m_0}) + \frac{\hbar}{m_0}\mathbf{k}\cdot\mathbf{P}_{lm} + \Delta_{lm} \right) c_m(\mathbf{k}) = E_l(\mathbf{k}) c_l(\mathbf{k}) \tag{2.2}$$

where the band offsets $E_m(0)$ enter, and

$$\mathbf{P}_{mn} = \langle m,0|\mathbf{p} + \frac{\hbar}{4m_0c^2}\boldsymbol{\sigma}\times\nabla V|n,0\rangle \tag{2.3}$$

$$\Delta_{mn} = \frac{\hbar}{4m_0^2c^2}\langle m,0|\mathbf{p}\cdot\boldsymbol{\sigma}\times\nabla V|n,0\rangle \tag{2.4}$$

are momentum and spin-orbit matrix elements evaluated in the basis for $\mathbf{k}_0 = 0$. Since $\mathbf{k}_0$ is a point of high symmetry, the transformation of the $|n,\mathbf{k}_0\rangle$ under the symmetry group of the crystal can be used to reduce the number of elements that have to be known.

The SO coupling in the momentum matrix elements $\mathbf{P}_{mn}$ can often be neglected. The SO-induced band splitting at the Γ point is still contained in Δ_{mn}. This way of treating SO interaction is useful since it reduces the number of parameters. In zinc blende structures, the matrix of the Δ_{mn} can be diagonalized by introducing basis functions of the T_d double group representations Γ_6, Γ_7 and Γ_8, leaving only the SO-induced gap at the Gamma point, Δ, as parameter. These basis functions are the lowest angular momentum eigenstates with $j = \frac{1}{2}$ and $j = \frac{3}{2}$.

The equation set (2.2) is infinite-dimensional and exact and contains only k-linear couplings between bands, if Δ_{lm} is diagonal. For practical purposes, second order quasi-degenerate perturbation theory (Appendix B) is used to construct a finite-dimensional, quadratic in k Hamiltonian. For example, in case one is interested in a model for only the Γ_6 band, which then should be well separated from the other bands, this leads to the effective mass approximation (n standing for Γ_6)

$$E_n(\mathbf{k}) = E_n(0) + \frac{\hbar^2}{2}\sum_{\alpha,\beta}^{x,y,z} k_\alpha \frac{1}{m^{\alpha\beta}} k_\beta \tag{2.5}$$

$$\frac{1}{m^{\alpha\beta}} = \frac{1}{m_0}\delta_{\alpha\beta} + \frac{2}{m_0^2}\sum_{m\neq n}\frac{P_{mn}^\alpha P_{nm}^\beta}{E_n(0) - E_m(0)}. \tag{2.6}$$

where the effective mass tensor of (2.6) contains corrections from other bands.

Thus it is possible to reduce the number of parameters to a finite number of effective masses and band offsets which may be determined experimentally, e.g. the band edge offsets by optical transitions and the effective masses from cyclotron resonance. Often, rather than calculating the parameters from band basis functions, the $\mathbf{k}\cdot\mathbf{p}$ method will be used to find a Hamiltonian including only a few parameters that can be fitted to experimental data.

2.2 Envelope function theory

The envelope function approximation (EFA) is a method for obtaining an effective Schrödinger equation for low-energy excitations in semiconductor systems where the lattice periodicity is

broken by a perturbation. Therefore, it is a generalization of $\mathbf{k} \cdot \mathbf{p}$ theory, which assumes a strictly periodic potential. Originally, the envelope function theory [Lut+55] was developed for systems where a potential varying slowly on the scale of a lattice constant adds up to the periodic crystal potential. Thus, the method may be applied to shallow impurity levels or magnetic fields [Lut+55], where the vector potential appearing by the Peierls substitution is space dependent, even if the magnetic field is constant.

In heterostructures, even when the chemical composition of the constituting semiconductors is similar, resulting in similar lattice constants and similar lattice-periodic basis functions, clearly, the perturbing potential is not slowly varying on the scale of the lattice constant. Instead, there will be a step-like behaviour at the interface of the constituents. However, still, the envelope function approximation has been widely and successfully applied to heterostructure problems (e.g. [And+87; Nov+05a]). Burt [Bur99] provides some arguments why this often works. Simply speaking, if the envelope function is slowly oscillating, it cannot "see" how abrupt the interface really is.

The basic recipe for obtaining an effective Schrödinger equation for the envelope function, is to take the $\mathbf{k}\cdot\mathbf{p}$ equations (2.2) or (2.5), and replace the band parameters, i.e. the effective masses and band edge potentials, by space dependent functions. For a heterostructure, one assumes that they are step functions. The momentum matrix elements $\mathbf{P}_{mn}$, on the other hand, are usually kept constant, since one assumes the lattice-periodic basis functions $u_{n,\mathbf{k}_0}(\mathbf{r})$ to be the same or at least very similar in all parts of the heterostructure. For the II-VI heterostructures where constituents both crystallize in zinc blende structure and have similar chemical composition, this assumption is well satisfied.

Further, a slowly perturbing potential may be added. Then, the crystal momentum $\mathbf{k}$ is replaced by the operator $\hat{\mathbf{k}} = -i\nabla$. Since $\hat{\mathbf{k}}$ no longer commutes with the space-dependent parameters, a symmetrization procedure must be applied, in order to obtain a Hermitian Hamiltonian in the resulting effective Schrödinger equation. There is an infinite number of possible symmetrizations. Different symmetrizations show different physical effects at heterostructure interfaces, so the choice is not arbitrary. However, since the $\mathbf{k} \cdot \mathbf{p}$ equations give no hint on interface effects, an *ad-hoc* symmetrization procedure is often used. Usually one enforces that for the Hamiltonian matrix H, $H_{ij}^\dagger = H_{ij}$ should hold, while Hermiticity just demands $H_{ij}^\dagger = H_{ji}$. In Section 2.3 we will discuss a simple special case of interest, where the condition $H_{ij}^\dagger = H_{ij}$ is required for mathematical consistency. The easiest and most common symmetrization is the substitution

$$\hat{k}_\alpha D_{ss'}^{\alpha\beta} \hat{k}_\beta \to \frac{1}{2}\left(\hat{k}_\alpha D_{ss'}^{\alpha\beta} \hat{k}_\beta + \hat{k}_\beta D_{ss'}^{\alpha\beta} \hat{k}_\alpha\right) \tag{2.7}$$

where $D_{ss'}^{\alpha\beta}$ is a generalization of an (inverse) effective mass term for the case of a multi-band model. We will refer to this formalism as *ad-hoc* symmetrized effective mass approximation (EMA). The question of symmetrization is equivalent to the question of choosing boundary conditions that connect the envelope function at both sides of the heterostructure [Win+93; For93].

Although the method seems very intuitive, a formal justification has long been missing. Thus, the quality of the approximation could only be estimated by comparison with exact calculations. Only in 1988, Burt [Bur88b] has given a formal derivation in form of his exact envelope equations. By applying several approximations, equations similar to the intuitive ad-hoc EMA may be obtained. Burt's theory brings about several advantages. First of all, the quality of the approximations may be understood, by analysing the neglected terms. Further, the ad-hoc

symmetrisation procedure is no longer needed. Instead, Burt's theory results in an Hermitian effective Hamiltonian with well-known order of the non-commuting operators. Thus, the Hamiltonian differs from the ad-hoc symmetrized one by a part containing antisymmetric operator combinations, and these terms may be physically relevant at the heterostructure interfaces.

In addition to continuity of the envelope function, one obtains other boundary conditions from integration of Burt's EMA equations or from current conservation. If effective masses in the heterostructure are different, the derivative of the envelope function becomes discontinuous. At first sight this seems to be in conflict to Burt's theory, which requires the envelope functions to be smooth. The conflict is resolved by recognizing that the EMA equations require a further approximation step, which removes fast oscillations in the smooth envelope function. The low-frequency part still has a smooth kink, which is the origin of the sharp kink (i.e. discontinuity of the derivative) that appears if one starts from the final EMA equations [Bur88a].

We will here sketch the derivation of Burt [Bur88b], to show the key approximations needed in order to arrive at the envelope function and EMA equations. We assume that the potential depends only on z-direction, corresponding to the growth direction of heterostructure, but this is no real restriction since the resulting equations are easily generalized to three dimensions. We also do not include SO interaction in the derivation. In the literature a generalization including SO terms can be found [Li+94]. The derivation contains additional complications, but after further approximations, the same EMA equations (2.15) are obtained, and only parameters become SO-dependent.

Let $\psi(z)$ be the exact solution of the Schrödinger equation

$$-\frac{\hbar^2}{2m_0}\frac{d^2\psi}{dz^2} + V(z)\psi = E\psi \tag{2.8}$$

where $V(z)$ is the microscopic potential which is however not really lattice periodic. Now one takes some - yet unspecified - complete basis $\{U_n\}_n$ of lattice periodic functions (lattice constant a). The envelope function components $F_n(z)$ are defined by the expansion

$$\psi = \sum_n F_n U_n \tag{2.9}$$

under the condition that the Fourier components of all the F_n are limited to the first Brillouin zone. This expansion is exact and gives a unique set of F_n, since ψ is unique. By this expansion, the vector $\mathbf{F}(z) = \{F_n(z)\}_n$ encodes the fast oscillations of $\psi(z)$ in its components for fixed coordinate z, while the slow oscillations are encoded in the change of $\mathbf{F}(z)$ with z. The imposed restriction to the Fourier expansion coefficients of $F_n(z)$, is the small but important difference to Bastard's [Bas88] envelope function approach. Bastard's envelope functions are simply assumed to be slowly oscillating, while Burt's envelope function components may oscillate rather fast, having only the upper frequency limit $\frac{\pi}{a}$.

The goal is to find expansions in the form of (2.9) for all terms of the Schrödinger equation, i.e. with Fourier components in the first BZ. Then, due to the uniqueness of the expansion, an equation for only the F_n may be written. This is easy for the kinetic term in (2.8), since taking derivatives of F_n does not give higher-oscillating components. Plugging in the expansion (2.9), we define momentum and kinetic terms similar to those occurring in $\mathbf{k}\cdot\mathbf{p}$ theory,

$$p_{mn} = \int \frac{dz}{a} U_m^* \left(-i\hbar\frac{dU_n}{dz}\right), \tag{2.10}$$

$$T_{mn} = \int \frac{dz}{a} U_m^* \left(-\frac{\hbar^2}{2m}\frac{d^2U_n}{dz^2}\right). \tag{2.11}$$

On the other hand, the term $V(z)\psi(z)$ requires more work since no restrictions apply to the oscillations of the factors. The Fourier transform of the potential term is $V_G(k) = \int \frac{dz}{L} V(z) e^{-i(k+G)z}$, with the sample length L, $|k| \leq \frac{\pi}{a}$ in the first BZ and $G = j\frac{2\pi}{a}$ some reciprocal lattice element. Fourier components of the U_n are defined by $U_{nG'} = \int \frac{dz}{a} U_n(z) e^{-iG'z}$. In the product $V(z)F_n(z)U_n(z)$, wave vectors k and G of $V(z)$ and k' of $F_n(z)$ and G' of $U_n(z)$ add up. The difficulty is that $k + k'$ no longer has to be in the first Brillouin zone. Therefore Burt defines k_1 and G_1 by $k + k' = k_1 + G_1$, with k_1 in the first BZ. Then, Burt's exact equation for the envelope components reads [Bur88a]

$$-\frac{\hbar^2}{2m_0}\frac{d^2F_n}{dz^2} - \frac{i\hbar}{m_0}\sum_m p_{nm}\frac{dF_m}{dz} + \sum_m \int dz' H_{nm}(z,z')F_m(z') = EF_n(z) \tag{2.12}$$

with $H_{nm}(z,z') = T_{nm}\delta(z-z') + V_{nm}(z,z')$ and

$$V_{nm}(z,z') = \frac{1}{L}\sum_{k,k'}\sum_{G,G'} U^*_{n,G+G_1} V_{G-G'}(k) U_{mG'} e^{ik_1 z - k'z'}. \tag{2.13}$$

In order to arrive at the commonly used effective mass equations, several approximations are made. First of all, we need to get rid of the integral in (2.12). This is done by splitting $V_{nm}(z,z')$ into local and non-local parts,

$$V_{nm}(z,z') = V_{nm}(z)\Delta(z-z') + V^{(nl)}_{nm}(z,z') \tag{2.14}$$

where we have defined $V_{nm}(z) = \int dz' V_{nm}(z,z')$, and $\Delta(z-z') = \frac{1}{L}\sum_{k'} e^{ik'(z-z')}$ is the delta function for functions with plane wave expansion limited to the first Brillouin zone. This separation in local and non-local parts is not unique but very practical. For $z \to z'$, $\Delta(z-z')$ drops to zero on the scale of the lattice constant. $V_{nm}(z)$ has a plane wave expansion limited to the first BZ, so dropping the non-local part eliminates the integral in (2.12). For $V(z)$ lattice-periodic, the non-local part vanishes. Further, in the situation where the perturbation $V(z)$ is slowly varying with highest non-zero Fourier component at k'_{max}, and the highest non-zero Fourier component of the envelope component $F_n(z)$ is at k_{max}, we can see that the non-local part in (2.12) vanishes identically if $k_{max} + k'_{max} < \frac{\pi}{a}$.

Burt analyses the local part $V_{nm}(z)$ for a heterostructure with a single interface at $z = 0$. At large distances left or right from the interface, $V_{nm}(z)$ takes the constant values $V^{(\pm)}_{nm}$. Then, assuming $V_{nm}(z) \approx V^{(-)}_{nm}(z)\theta(-z) + V^{(+)}_{nm}(z)\theta(z)$ is called abrupt step approximation (with $\theta(z > 0) = 1$ and $\theta(z \leq 0) = 0$). This approximation neglects Gibbs oscillations that exist in $V_{nm}(z)$. However, these are localized closely to the interface, and therefore, for slowly varying envelope function components, they cannot contribute much in (2.12).

Up to now, the basis of lattice-periodic functions U_n has been left unspecified, which is clearly an advantage in the heterostructure problem since in general, there will be no basis that diagonalizes $V^{(+)}_{nm}$ and $V^{(-)}_{nm}$ at the same time. But the assumption that the same basis of band-edge Bloch functions may be used for writing the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian of all parts of the heterostructure, is omnipresent in envelope function theory. For practical purposes, where the heterostructure constituents possess the same crystal symmetry, the assumption that there is a common basis which approximately diagonalizes $V^{(+)}_{nm}$ and $V^{(-)}_{nm}$, is well satisfied. In this basis, we have $V_{nn}(z) \approx E_n(z)\delta_{nm}$. In the bulk of the structure (i.e. far from the heterostructure interface), this already reduces (2.12) to the usual $\mathbf{k}\cdot\mathbf{p}$ equations (2.2) with appropriate band edge potentials.

It is common to apply the EFA to heterostructures, where any general perturbation $\tilde{V}(z)$ that breaks lattice periodicity, can in no way be considered as slowly varying or small. Note that $V(z)$ includes a common periodic potential $V_{per}(z)$ of the cores, and here we are talking about $\tilde{V}(z) = V(z) - V_{per}(z)$. Burt's derivation shows that even if $\tilde{V}(z)$ is not slowly varying, it suffices that the envelope function components are slowly varying.

In order to obtain effective mass equations, one has to apply quasi-degenerate perturbation theory (Appendix B) just as in the case of $\mathbf{k}\cdot\mathbf{p}$ theory. We denote the set of bands that we wish to include exactly, with indices s, s'. The other bands, which should be well separated from the former in energy, are labelled with indices l, l'. Treating coupling to the remote bands as perturbation, second order perturbation theory yields the effective mass equations. In the form generalized to 3D coordinates, they read [Bur88b; Pfe00]

$$\sum_{s'} \left(\sum_{\alpha,\beta}^{x,y,z} \hat{k}_\alpha D^{\alpha\beta}_{s,s'} \hat{k}_\beta + \frac{\hbar}{m_0} \sum_{\alpha}^{x,y,z} p^\alpha_{ss'} \hat{k}_\alpha + \delta_{ss'} E_s(\mathbf{r}) \right) F_{s'}(\mathbf{r}) = E F_s(\mathbf{r}) \tag{2.15}$$

where the space-dependent (inverse) effective mass tensor is given by

$$D^{\alpha\beta}_{ss'} = \frac{\hbar^2}{2m_0} \left(\delta_{ss'}\delta_{\alpha\beta} + \frac{2}{m_0} \sum_{l \neq \{s,s'\}} \frac{p^\alpha_{sl} p^\beta_{ls'}}{E - E_l(\mathbf{r})} \right). \tag{2.16}$$

The energy dependency of this tensor can be lifted by assuming $E \approx E_s(\mathbf{r})$. Compared to the ad-hoc symmetrization (2.7), the ordering of the non-commuting operators $\hat{k}_\alpha$, $D^{\alpha\beta}_{ss'}$, $\hat{k}_\beta$ is clearly given in (2.15). Note[1] that $[p^\alpha_{s,s'}, \hat{k}_\alpha] = 0$ because a globally unique basis has been used for evaluation of the momentum matrix elements $p^\alpha_{s,s'}$. Since we may write

$$\hat{k}_\alpha D^{\alpha\beta}_{ss'} \hat{k}_\beta = \frac{1}{2}\left(\hat{k}_\alpha D^{\alpha\beta}_{ss'} \hat{k}_\beta + \hat{k}_\beta D^{\alpha\beta}_{ss'} \hat{k}_\alpha \right) + \frac{1}{2}\left(\hat{k}_\alpha D^{\alpha\beta}_{ss'} \hat{k}_\beta - \hat{k}_\beta D^{\alpha\beta}_{ss'} \hat{k}_\alpha \right), \tag{2.17}$$

we see that there are antisymmetric terms which would be removed by the ad-hoc symmetrization, although they can be important for interface effects in heterostructures. It is also possible to generalize the result (2.15) by including a magnetic field via the substitution $\hat{\mathbf{k}} \to -i\nabla + \frac{e}{\hbar}\mathbf{A}$ with the vector potential $\mathbf{A}$ corresponding to a magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$, and $-e < 0$ is the electron charge. In this case, the antisymmetric terms are proportional to the magnetic field.

In the derivation shown above, we have not included spin-orbit terms. There are two ways to extend the derivation for that. Firstly, one may take a band basis that does not depend on spin, and include spin-orbit coupling as perturbation [Lut+55; Pfe00]. Just like in $\mathbf{k}\cdot\mathbf{p}$ theory, matrix elements $\Delta_{ss'}$ of the perturbation will give a correction to the band edges depending on spin and band indices, and it may also depend on the coordinate $\mathbf{r}$, if SO coupling changes in the heterostructure. The correction can be diagonalized by a basis transformation, and the new basis will correspond to the double group representations of the underlying Bravais lattice. Usually, these representations are used to label the bands.

Secondly, one may start from a spin-dependent band basis adjusted to the double group representations [Li+94] and develop Burt's theory with SO terms included from the beginning. This leads to some extra complications, but in the end the same result (2.15) can be obtained, where the spin-orbit coupling renormalizes the effective band parameters.

[1] More precisely, this holds only if SO coupling is neglected in the momentum matrix elements or taken as the same in all parts of the heterostructure, compare Eq. (2.3).

2.3 Symmetrization of k-linear terms in effective models

Up to now, we have not discussed the question of symmetrizations for terms linear in momentum. This was not necessary, since we argued that the lattice-periodic basis needed for the development of the envelope funtion theory, will be very similar in all parts of the heterostructure, so the momentum matrix elements should not be space dependent. In particular, in Burt's envelope function approach, the momentum matrix elements are constants. In a situation with SO coupling the Kane $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian [Kan57] can contain further terms linear in $\mathbf{k}$, related to broken spatial inversion symmetry. Most importantly, effective perturbative models for 2D systems with broken inversion symmetry [Win05] contain $\mathbf{k}$-linear terms for Rashba SO coupling, and their value may strongly depend on the coordinate. In some situations where the momentum matrix element is not truly constant, it has been proposed to build effective models that enforce it to stay constant, but instead, adjust the effective masses of the relevant $\mathbf{k}\cdot\mathbf{p}$ model, in order to find a good approximation to a given band structure (compare discussion in [For97]).

In Chapter 10 we will need an effective Hamiltonian for a two-dimensional system with space dependent Rashba SO coupling, which is linear in $\mathbf{k}$. It turns out that in this case, the usual spin-independent symmetrization procedure

$$\alpha(x)\hat{k}_x\sigma_y \to \frac{1}{2}\sigma_y(\alpha(x)\hat{k}_x + \hat{k}_x\alpha(x)) \tag{2.18}$$

has to be used. One could imagine using a symmetrization procedure that makes use of the spin space to obtain other Hermitian Rashba Hamiltonians, e.g.

$$\alpha(x)\hat{k}_x\sigma_y \to \begin{pmatrix} & -i\alpha(x)\hat{k}_x \\ i\hat{k}_x\alpha(x) & \end{pmatrix}. \tag{2.19}$$

Interestingly, such a choice of the Hamiltonian is mathematically inconsistent with the assumption of a continuous envelope function, when the kinetic energy term is added. We will show this below. Let's consider the prototype 2-band Hamiltonian

$$H = \begin{pmatrix} E_c + \hat{k}_x A_c \hat{k}_x & i(\hat{k}_x Q + P\hat{k}_x) \\ -i(\hat{k}_x P + Q\hat{k}_x) & E_v + \hat{k}_x A_v \hat{k}_x \end{pmatrix}. \tag{2.20}$$

For arbitrary space-dependent parameter functions, this Hamiltonian is Hermitian. Our goal is to analyse whether step-like variations of the parameters lead to consistent boundary conditions. If this is not the case, the Hamiltonian must be considered invalid. We assume the wave function $\psi = (\psi_1, \psi_2)$ to be continuous. In fact, ψ should be understood as 2-component envelope function. The parameters P, Q, E_c, and E_v are assumed to be step functions with steps at $x = 0$. Integration of $H\psi = E\psi$ over the interval $[-\epsilon, \epsilon]$ gives

$$\mathcal{O}(\epsilon) + \left[-A_c\psi_1' + Q\psi_2\right]_{-\epsilon}^{\epsilon} = 0 \tag{2.21}$$

$$\mathcal{O}(\epsilon) + \left[-P\psi_1 - A_v\psi_2'\right]_{-\epsilon}^{\epsilon} = 0, \tag{2.22}$$

where we used the continuity of ψ to find $\int_{-\epsilon}^{\epsilon} dx\, P\psi_2' = \mathcal{O}(\epsilon)$ and $\int_{-\epsilon}^{\epsilon} dx\, Q\psi_1' = \mathcal{O}(\epsilon)$.

Let's compare with the continuity of $\hat{V}\psi$, where

$$\hat{V} = \frac{\partial H}{\partial \hat{k}_x} = \begin{pmatrix} \{A_c, \hat{k}_x\} & i(Q+P) \\ -i(P+Q) & \{A_v, \hat{k}_x\} \end{pmatrix}. \tag{2.23}$$

The current $j(x)$ should be continuous, and it is given by

$$j(x) = \psi^\dagger \hat{V} \psi = \psi_1^* \{A_c, \hat{k}_x\} \psi_1 + \psi_2^* \{A_v, \hat{k}_x\} \psi_2 + (\psi_1^* \psi_2 - \psi_2^* \psi_1) i(Q+P). \tag{2.24}$$

On the other hand, from the boundary conditions (2.21),(2.22) we obtain that the following function is continuous:

$$-iA_c \psi_1^* \psi_1' - iA_v \psi_2^* \psi_2 + iQ\psi_1^* \psi_2 - iP\psi_2^* \psi_1. \tag{2.25}$$

If A_c and A_v are constant, $\{A_{c/v}, \hat{k}_x\} = 2A_{c/v}\hat{k}_x$. If additionally $P = Q$, continuity of (2.25) is equivalent to the continuity of $j(x)$. In that case, the boundary conditions (2.21), (2.22) also give the continuity of $\hat{V}\psi$.

Subtracting (2.24) from 2 times (2.25), gives the continuous expression

$$i(P-Q)(\psi_1^* \psi_2 + \psi_2^* \psi_1). \tag{2.26}$$

Since $\psi_1^* \psi_2 + \psi_2^* \psi_1$ is assumed to be continuous and is in general non-zero, $P - Q$ should also be continuous. If we write $P = \alpha\theta(x) + p$, $Q = \beta\theta(x) + q$ with the Heaviside function $\theta(x)$, we see that we must have $\alpha = \beta$. For the constants p, q one can set $p = q$ without loss of generality, and we conclude that $P = Q$ is the only reasonable choice in the Hamiltonian (2.20), if quadratic terms are present ($A_{c/v} = \text{const} \neq 0$).

Note that our result, that the usual symmetrization (2.18) should be used, is a special case and should not be taken as general result for other Hamiltonians. In particular, the assumption $A_{c/v} \neq 0$ is important. If $A_c = 0$, taking the limit $\epsilon \to 0$ in (2.21), shows that $Q\psi_2$ will be continuous. But since Q is a step function, ψ_2 must be a step function, in contrast to our assumption (unless $\psi_2(0) = 0$). A problem similar to the case $A_c = 0$ has been discussed by Foreman [For97] in the context of an 8-band $\mathbf{k} \cdot \mathbf{p}$ model where the momentum matrix element is space-dependent. In Foreman's case, putting $A_c = 0$ is used as a trick to remove spurious large-k solutions, and it is compensated for by adjusting the values of the k-linear terms, in order to keep a good approximation to the desired band structure. The mentioned contradiction in the limit $A_c \to 0$ leads Foreman to the conclusion that one should choose $Q = 0$, i.e. that a non-trivial symmetrization must be used in the case $A_c = 0$. In another analysis by Foreman [For93] where Burt's theory is applied to a valence band model including SO coupling, the correct Hamiltonian is also *not* of the symmetrized form. Again, the analysed Hamiltonian in [For93] is not comparable to our simple prototype Hamiltonian (2.20).

Other symmetrizations

For a product of non-commuting operators like $\alpha(x)\hat{k}_x$, there are infinitely many possible ways of symmetrization, and if $\alpha(x)$ is given by a step function, the physical difference of these corresponds to different choices of boundary conditions [Win+93]. In a phenomenological approach where the operator ordering of an effective model is not known from first principle calculations, a simple picture that explains the physical difference of the symmetrizations will be useful. The most commonly used symmetrization are

$$H^{(a)} = \frac{1}{2}\{\alpha(x), \hat{k}\}, \quad H^{(b)} = \sqrt{\alpha(x)}\hat{k}\sqrt{\alpha(x)}. \tag{2.27}$$

The question about the physical difference of $H^{(a)}$ and $H^{(b)}$ is easiest answered by looking at corresponding discretized, real-space matrix representations on a lattice with discrete coordinates

x_i. We use the substitutions $\alpha(x) \to \delta_{ij}\alpha(x_i)$ and $\hat{k} \to t_{ij} = t(x_i - x_j)$ with a function $t(a) = \frac{1}{2ai} = -t(-a)$ and $t(\Delta x) = 0$ otherwise, and find the discrete representations

$$H_{ij}^{(a)} = \frac{\alpha(x_i) + \alpha(x_j)}{2} t_{ij}, \quad H_{ij}^{(b)} = \sqrt{\alpha(x_i)\alpha(x_j)}\, t_{ij}. \tag{2.28}$$

For a step function $\alpha(x_i) = \alpha\theta(x_i)$ modelling an interface at $x = 0$, the geometrical average in $H^{(b)}$ corresponds to a sharper interface, and $H_{ij}^{(b)} = \alpha(\min(x_i, x_j))t_{ij}$. In this case we have $H_{ij}^{(a)} = (H_{i-1,j-1}^{(b)} + H_{ij}^{(b)})/2$, showing that $H^{(a)}$ is the average of two models of the type $H^{(b)}$, shifted by one lattice site.

2.4 Wave matching and lattice models for envelope functions

The method of solving the Schrödinger equation for different regions, where the value of a potential is different but constant in each region, and then obtaining a global solution by finding appropriate coefficients to match the solutions at interfaces, is known from basic quantum mechanics courses and allows to treat simple transport problems analytically, like the calculation of transmission and reflection probabilities in simple heterostructures. For a second order differential equation, in addition to continuity of the wave function, a second condition is obtained by integration over the interface. This condition is related to current conservation and in the simplest case, it corresponds to the continuity of the derivative of the wave function. In the general case however, the latter may not be assumed.

In the last section we have already seen that even for small, k-linear models for envelope functions, difficulties with symmetrization if the precise operator ordering is not known from Burt's envelope function theory, and even mathematical inconsistencies can appear in wave matching calculations. Even worse, often effective models derived by perturbation theory contain higher orders of k. For such higher order differential equations, it is very difficult to obtain physically meaningful boundary conditions at the heterostructure interface [Li+07]. The envelope function method naturally results in models of high matrix dimension, and due to increased mathematical complexity, equal problems may appear for large systems of linear or quadratic differential equations [Kis+98]. An unfortunate situation would be if boundary conditions heavily depended on the approximation used during a derivation, but have large influence on interesting solutions, if those are localized at the interface.

To illustrate some of these problem within a simple model, let us consider an envelope function approximation model for heavy holes in a 2D system with cubic Rashba SO coupling and an interface at $x = 0$,

$$\left(\frac{\hbar^2 k^2}{2m} + \frac{1}{2}\begin{pmatrix} & \{\alpha(x), ik_-^3\} \\ \{\alpha(x), -ik_+^3\} & \end{pmatrix}\right)\psi = E_F\,\psi \tag{2.29}$$

where the Rashba coupling $\alpha(x)$ is assumed to be a step function, $\alpha(x) = \alpha_-$ for $x < 0$ and $\alpha(x) = \alpha_+$ for $x > 0$. In the bulk regimes, we find the eigenenergies $E_\pm = \frac{\hbar^2 k^2}{2m} \pm \alpha k^3$ and corresponding eigenmodes $\psi_\pm = \frac{1}{\sqrt{2}}(\pm i\frac{k_-^3}{k^3}, 1)^T$. The dispersion is plotted in Fig. 2.1a. We see that in the interesting regime of energies, the Fermi energy E_F has six intersections with positive and negative values of k, independent of the direction of $\mathbf{k}$. For simplicity, let us assume that the incoming wave is perpendicular to the interface, otherwise we will have to care about simultaneous conservation of energy and k_y in order to find solutions. The intersections $\{k_l\}_{l=1,..,6}$

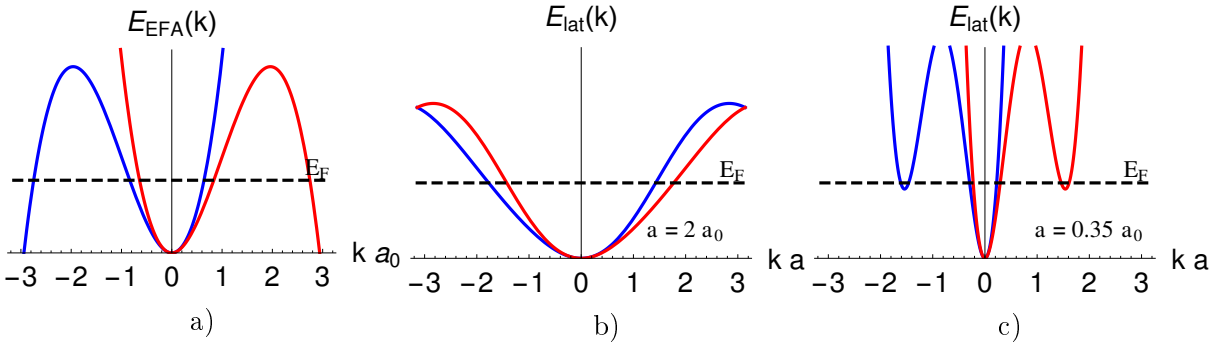


Figure 2.1: Bulk energy dispersion of the cubic Rashba model (2.29). To illustrate the problem of spurious solutions and how they are eliminated in a lattice model, we choose a Rashba SO coupling $\alpha = 0.34\frac{\hbar^2}{2ma_0}$. In a), the dispersion of the unmodified EFA model is shown, which should be valid in the regime $|k|a_0 < 1$. Here a_0 is the lattice constant of the original physical system. There are six intersections with the Fermi energy, four of them are physical and two are spurious. In b), the dispersion is shown for a lattice model of the discretized EFA model, with a new lattice constant $a = 2a_0$. The dotted line shows the same Fermi level as in a). There are no spurious solutions. In c) the dispersion of the lattice model with $a = 0.35a_0$ is shown. For $a \leq 0.35a_0$, the problem of Fermion doubling appears and there are four spurious solutions.

will be needed to construct the global envelope wave function ψ from the corresponding modes, including transmitted and reflected parts, by the ansatz

$$\psi = \psi_{in} e^{ik_{in}x} + \sum_{l=1}^{6} c_l\, \theta(x v_l)\, \psi_l e^{ik_l x}, \tag{2.30}$$

where ψ_{in} is an incoming mode with wave vector k_{in}, the ψ_l are the transmitted and reflected modes which should be evaluated with parameters of the right and left bulk regions, respectively, c_l are the corresponding transmission and reflection amplitudes, and the Heaviside θ function is used to select modes of the correct direction of propagation, by the sign of their velocity v_l and the side of the interface.

However, we know that the two largest positive and negative wave vectors must be regarded as spurious (unphysical) solutions[2], since they appear only because the model (2.29) is obtained from the fit of a cubic dispersion to the correct dispersion, and this fit is correct only for small k. By this, we reduce the number of reflection and transmission coefficients that need to be calculated, to four.

This nicely matches the set of four equations that we obtain from the conditions of continuity of the spinor ψ and $\frac{\partial H}{\hbar \partial k_x}\psi$ at the interface,

$$\psi|_{x=0^-} = \psi|_{x=0^+} \tag{2.31}$$

$$\frac{1}{\hbar}\frac{\partial H}{\partial k_x}\psi\bigg|_{x=0^-} = \frac{1}{\hbar}\frac{\partial H}{\partial k_x}\psi\bigg|_{x=0^+} \tag{2.32}$$

So the problem seems to be solvable without large complications. However, the condition (2.32) that we have put by the intuition of current conservation, is not mathematically correct! We

[2] Problems with spurious solutions are omnipresent in applications of envelope function theory, see e.g. [Sch+85]

only know that this would be the mathematically correct condition in the linear Rashba model. In Appendix (E) a derivation of a generalized continuity equation is shown, which proves that the current density operator for a Hamiltonian H including linear and quadratic terms in k can be written as $\mathbf{J}(\mathbf{r}) = \frac{1}{2}\left\{\frac{\partial H}{\partial \mathbf{k}}, \delta(\mathbf{r}-\hat{\mathbf{r}})\right\}$, corresponding to matching condition (2.32). By contrast, for a cubic model like (2.29), this is not the precise mathematical condition, which would be obtained from integration over the Schrödinger equation at the boundary [Li+07]. A consistent, mathematical precise treatment of the third order differential equation (2.29) needs more boundary conditions and would require all six solutions k_l, including the spurious ones. In spite of being not mathematically correct, we can expect that our simple calculation based on the matching conditions (2.31) and (2.32) will give a better approximation to the correct physical behaviour than the mathematical consistent solution, which includes spurious solutions.

Instead of proving this statement, we want to show a different method to avoid problems with spurious solutions, which is similar to an idea by Winkler [Win+93]. Since the regime of validity of the EFA is limited to small k, our idea is to discretize the effective multiband Hamiltonian on a lattice, keeping a good approximation only for small k. The lattice constant a may be very different (normally larger) than a_0 of the original Bravais lattice of the physical system. The Brillouin zone of the artificial lattice model should be small enough to cut off any spurious solutions. By using the translation operator in the way $e^{i\hat{k}_x a}\psi(x) = \psi(x+a)$, we obtain operator expressions corresponding to the discretizations of $\hat{k}_i$ and $\hat{k}_i^2$ on a lattice,

$$\hat{k}_x\psi(x) = \frac{1}{i}\partial_x\psi(x) \to \frac{1}{2ia}\left(\psi(x+a) - \psi(x-a)\right) = \frac{1}{a}\sin(\hat{k}_x a)\psi(x) \tag{2.33}$$

$$\hat{k}_x^2\psi(x) = -\partial_x^2\psi(x) \to \frac{1}{a^2}\left(2\psi(x) - \psi(x+a) - \psi(x-a)\right) = \frac{1}{a^2}(2 - 2\cos(\hat{k}_x a))\psi(x) \tag{2.34}$$

and analogue for other components of $\hat{\mathbf{k}}$. If a magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ has been included in the effective model by the Peierls substitution $\hat{k}_i \to \hat{q}_i = \hat{k}_i + \frac{e}{\hbar}A_i$, the discretization procedure has to be changed a bit, in order to obtain a gauge covariant lattice model. Using the Landau gauge $\mathbf{A} = \frac{1}{2}\mathbf{B} \times \mathbf{r}$ for a homogeneous field $\mathbf{B}$, we have $[\hat{k}_i, A_i] = 0$, and a substitution similar to (2.33) can be used to find

$$\hat{q}_x\psi(x) \to \frac{1}{a}\sin(\hat{q}_x a)\psi(x) = \frac{1}{2ia}\left(e^{iaeA_x/\hbar}\psi(x+a) - e^{-iaeA_x/\hbar}\psi(x-a)\right), \tag{2.35}$$

and analogue for other components $\hat{q}_i$. Compared to (2.33), including the vector potential results in extra phase factors. This result may be generalized to find a general substitution rule that is equivalent to the Peierls substitution, to include the vector in the lattice model in a gauge covariant way [Dat07; Rot09]. If the lattice model without the vector potential is given by the Hamiltonian matrix $H_{ij} = \langle \mathbf{r}_i | H | \mathbf{r}_j \rangle$, where we have denoted lattice coordinates as $\mathbf{r}_i$, we have to add phase factors to this matrix, by the substitution

$$H_{ij} \to H_{ij} e^{i\frac{e}{\hbar}\int_{\mathbf{r}_i}^{\mathbf{r}_j} d\mathbf{r}\cdot\mathbf{A}(\mathbf{r})}. \tag{2.36}$$

In Figure 2.1b, the dispersion relation of a lattice model corresponding to the cubic Rashba problem (2.29) is plotted. We can see that for a reasonable choice of the lattice constant a, the spurious solutions disappear, while the interesting regime of the dispersion relation is well approximated by the lattice model. The Green's function formalism of Chapter 5 can be used to implement the wave matching calculation, and there are no ambiguities concerning boundary

conditions or current conservation. The method is explained in more detail in Appendix F and will be applied in Chapter 10.

Here, we want to emphasize that the lattice-based solution comes at a price, and can require a careful adjustment of the lattice constant a. Since the discretization re-introduces periodicity of the dispersion relation $E(k)$, there may be problems if k-linear terms are dominant in the Hamiltonian, like in effective models for graphene. Let us assume the most simple dispersion of an effective model, $E(k) = \alpha k$, so there is a single intersection with the Fermi level, giving a single k_F. The lattice regularization is $E_{lat}(k) = \frac{\alpha}{a}\sin(ak) = \frac{\alpha}{a}\sin(\pi - ak)$, which crosses the Fermi level two times. This phenomenon is called Fermion doubling. Generally in d dimensions, the lattice equivalent of the linear Dirac dispersion will show 2^d-fold the number of original Fermions [Sta82]. Since k is assumed to be small, the extra solutions are highly oscillating. The Fermion doubling problem does not apply for quadratic-in-k Hamiltonians, since the number of crossings with E_F is the same in the original and the lattice model. Generally speaking, the lattice regularization doubles the set of solutions at the Fermi energy only for odd powers of operators $\hat{k}_i$. If the Hamiltonian contains both even and odd powers of $\hat{\mathbf{k}}$, the appearance of highly oscillating spurious solutions will depend on band parameters and Fermi energy. Figure 2.1c shows the lattice dispersion of the cubic Rashba model that we have discussed before, with a lattice constant a that is small enough to generate the spurious solutions due to Fermion doubling. Luckily, highly oscillating extra solutions are easily detected during the wave matching procedure. The problem is very relevant to electronic transport in graphene [Two+08], and if disorder is included in the lattice model, scattering between low k and high k may prohibit the sorting out of unphysical results. Then, it will be preferable to simulate the atomic lattice of graphene instead of the effective Dirac model.

To finalize the discussion of Fermion doubling, let us emphasize that it can also be a physical feature and not merely a numerical nuisance. Let us consider a physical 1D lattice, i.e. a chain. Periodicity of the energy dispersion ensures that for a propagating mode of one direction, another mode propagating in the opposite direction exists. For this reason (in equilibrium, but regardless of time reversal symmetry) it is not possible to build a wire that conducts in only one direction. However, almost current-rectifying effects are possible in topologically non-trivial systems, where current-carrying states are localized at the sample edges, thus making it possible to select the direction of propagation by selecting the sample side.

To conclude, the wave matching calculation for higher-order models like e.g. the cubic model (2.29) can be performed on a lattice in order to remove spurious solutions, however the lattice constant must be chosen carefully to avoid Fermion doubling. For moderate spin-orbit coupling, $a = a_0$ or $a = 2a_0$ does the job, compare Fig. 2.1b.

Chapter 3

Theory of invariants

The method of invariants provides a systematic formalism to derive all possible Hamiltonians of a system that are allowed by symmetry - e.g. the symmetry of a semiconducting crystal. First, one needs to specify the basis in which the Hamiltonian should be written. Usually, this will be a set of bands close to the energy of interest, i.e. the Fermi energy. Then, all possible terms up to some order in $\mathbf{k}$, possibly also including external fields, can be written down and be classified by their transformation under the point group of the crystal. The method is very useful in combination with the $\mathbf{k}\cdot\mathbf{p}$ or envelope function method. While the latter also give expressions for energy offsets or other band parameters, the method of invariants can only provide information about the degeneracy of bands and the analytic form of the Hamiltonian. In practice, also the expressions obtained with the $\mathbf{k}\cdot\mathbf{p}$ method need to be adjusted to experimental data, since ab-initio calculations are very difficult. Therefore, both methods are of comparable use. The method of invariants is well established and is broadly discussed e.g. in the comprehensive works [Pik+74] and [Dre+08]. Also, a short introduction can be found in [Win05]. First, we will give a short overview of the method. We will use the Hamiltonian of a spin-$\frac{3}{2}$ in a quadrupole field as example of how to apply the method. The obtained Hamiltonian will be the basis of the discussion in Chapter 8. Later, we will also make use of the method of invariants in the derivation of an effective Hamiltonian for a HgTe/CdTe QW.

The group of all the operations $\hat{g}$ that leave the system invariant, forms the symmetry group G of the system. Each symmetry element commutes with the Hamiltonian, $[H, \hat{g}] = 0$. In the following, we use the symbol g for the abstract mathematical group element corresponding to the operation $\hat{g}$. In the case of the crystal, group operations consist of lattice translations and rotation or reflection operations, where the latter are called point group operations, and they form a subgroup of the orthogonal group $O(3)$. We will write $\mathcal{D}(g)$ for the unitary representation of group element g on the matrix space of the Hamiltonian H, so $\mathcal{D}^{-1}(g) = \mathcal{D}^{\dagger}(g) = \mathcal{D}(g^{-1})$. If the Hamiltonian matrix depends on a vector, which could be e.g. the direction of an external magnetic or electric field, or simply the momentum $\mathbf{k}$, the point group operation also acts on this vector. However, the inverse operation is needed for the vector, regarding the operation as transformation of the coordinate system. We use the generic symbol $\mathcal{K}$ to represent all contained vectors, and write $g^{-1}\mathcal{K}$ for the set of rotated vectors. So the condition of invariance $[\hat{g}, H] = 0$ can be written as [Win05]

$$\mathcal{D}(g)H(g^{-1}\mathcal{K})\mathcal{D}^{-1}(g) = H(\mathcal{K}). \tag{3.1}$$

Since in the Bloch basis, the translation by a lattice vector $\mathbf{R}$ is just a complex number $e^{i\mathbf{kR}}$ and

cancels in (3.1), we need to consider only representations of the point group in the following.

A representation $\mathcal{D}(g)$ is called reducible if there is a non-trivial block-diagonalization of $\mathcal{D}(g)$ that works for all (point) group elements g at the same time. Otherwise, it is called irreducible. Since each irreducible block of $\hat{g}$ commutes with H, the dimension of the block corresponds to the degeneracy of an energy eigenstate or band.

The orthogonal group $O(3) = SO(3) \times C_i$ can be written as direct product of proper rotations and the group C_i that consists of the identity and the parity (space inversion) operator, since the latter commutes with all other elements. Here we are mainly interested in the continuous part $SO(3)$ of the group, since it corresponds to the $SU(2)$ representations. The inversion element is not contained in $SU(2)$ and should not be confused with the 2π rotation, if this evaluates to -1. Irreducible representations of $SO(3)$ are also irreducible representations of $O(3)$[1]. Irreducible $O(3)$ representations are denoted by $\mathcal{D}^{(j)\pm}$ with an additional superscript for the transformation under parity. The parity operator is also important for symmetry considerations and gives additional restrictions to the allowed terms, but transformation under parity can be analysed separately. Usually, the terms bonding and antibonding refer to the different parities, and they are often denoted with a superscript like in $\mathcal{D}^{\pm}(g)$.

The group relevant for rotation of a spinor is actually $SU(2)$ and not $SO(3)$. $SU(2)$ covers $SO(3)$ twice. Geometrically, this can be seen from the fact that a 2π rotation takes a vector to minus itself in half-integer representations of $SU(2)$, and a 4π rotation is needed to obtain the identity, while an operation that changes the sign of a vector is not contained in $SO(3)$.

The needed irreducible representations are the well-known representations $\mathcal{D}^{(j)}(g) = e^{i\boldsymbol{\omega}\mathbf{J}^{(j)}}$ corresponding to integer or half-integer angular momentum j, where the vector of angles $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ parametrizes the element g, and the components of the vector $\mathbf{J}^{(j)} = (J_x^{(j)}, J_y^{(j)}, J_z^{(j)})$ are $(2j+1)$-dimensional matrices that satisfy the well-known commutation relations of angular momentum j,

$$[J_n^{(j)}, J_m^{(j)}] = ij\epsilon_{nmk} J_k^{(j)}. \tag{3.2}$$

Explicitly, they are given by the definitions $J_+^{(j)}|j,m\rangle = \sqrt{(j-m)(j+m+1)}|j,m+1\rangle$, $J_-^{(j)} = J_+^{(j)\dagger}$, $J_x^{(j)} = \frac{1}{2}(J_+^{(j)} + J_-^{(j)})$, $J_y^{(j)} = \frac{1}{2i}(J_+^{(j)} - J_-^{(j)})$ and $J_z^{(j)} = \mathrm{diag}(-j, -j+1, ..., j)$.

Only for the half-integer representations of $SU(2)$, a 2π rotation gives the group element $-\mathbb{1}$, while for integer representations, a 2π rotation gives the identity. Thus, integer representations of $SU(2)$ are not faithful and are really representations of $SO(3)$ (the term *faithful* refers to a one-to-one correspondence of group elements and their representations, whereas integer representations have a two-to-one correspondence). The half-integer representations are also called *double group* representations. The 2π rotation is a group element that commutes with all other elements. Therefore, compared to the number of group elements in $SO(3)$ (or a discrete subgroup thereof), we can say that there are twice as many elements in the $SU(2)$ representation.

Finally, we also need to mention that the point group of the crystal is just a discrete subset of the full orthogonal group $O(3)$ of arbitrary rotations and reflections. The group T_d of a tetrahedron is most relevant to us, since it is the point group of the zinc blende structure. When embedded in a cube, the tetrahedron connects every second corner of the cube. Thus, T_d does not contain the spatial inversion. T_d has a total of 12 elements, which are grouped into five classes [Pik+74],

$$(e), \quad (4c_3, 4c_3^2), \quad (3c_2), \quad (3s_4, 3s_4^3), \quad (6\sigma). \tag{3.3}$$

[1] This statement holds for the continuous groups, but not necessarily for discrete subgroups of $O(3)$ vs. $SO(3)$.

Here e is the identity, c_3 are threefold and c_2 twofold rotations, s_4 are rotations about $\pi/4$ combined with inversion, and σ are reflections. In finite groups, the number of irreducible representations is equal the number of classes. For T_d, they are denoted as $\Gamma_1, ..., \Gamma_5$ [2] [3]. On the other hand, the group of rotations of a cube is called O. It also has 12 elements, and is mathematically isomorphic to T_d. Therefore, the irreducible representations are the same as for T_d, but their physical meaning is different, since there are no reflections in O. In the group tables (e.g. [Pik+74], Table 11.1), the difference is seen in the basis functions for T_d and O, respectively. The vector (x, y, z) transforms like Γ_5 in T_d, but like Γ_4 in O. This can lead to some confusion when applying the method of invariants to a T_d system, especially since it is common to find approximate results for a zinc blende crystal from the diamond crystal, which has the point group O_h. $O_h = O \times C_i$ is the group of a cube including reflections and spatial inversion, and it has 24 elements and 10 irreducible representations $\Gamma^{\pm}_{1,...,5}$, which carry a parity superscript. Physically, the even/odd parity corresponds to bands of bonding/antibonding character in the crystal with diamond structure. The O_h approximation to T_d will exclude the bulk inversion asymmetry (BIA) terms.

Since T_d contains reflection operations, and reflections can always be rewritten as rotation times inversion, one is tempted to (imprecisely) classify the T_d representations according to their parity. Looking at Table 3.1, it seems that Γ_1, Γ_4 are parity-even and Γ_2, Γ_5 are parity-odd. Then, Γ_1 should transform as scalar, Γ_2 as pseudoscalar (e.g. $\mathbf{k} \cdot \mathbf{B}$), Γ_4 as pseudovector (e.g. $\mathbf{J}$ and $\mathbf{B}$) and Γ_5 as vector (e.g. $\mathbf{k}$ and the electric field $\mathbf{E}$). However, this does not hold strictly, since there are both parity-even tensors transforming like Γ_5, e.g. $(k_y k_z, k_z k_x, k_x k_y)$, and parity-odd tensors of higher order transforming like Γ_4. By contrast, the representations of $O_h = T_d \times C_i$ may be truly classified into even/odd transformation under parity.

In discrete point groups, there will still be representations where a 2π rotation is non-trivial, and those are still called double group representations. The other representations, where a 2π rotation is identity, are called single group or $SO(3)$ representations. In spin-orbit coupled systems, we need the double group representations of T_d, which are formed by taking the product with the spinor representation $\mathcal{D}^{(1/2)+} = \Gamma_6$, as

$$\Gamma_1 \otimes \Gamma_6 = \Gamma_6, \quad \Gamma_2 \otimes \Gamma_6 = \Gamma_7, \quad \Gamma_3 \otimes \Gamma_6 = \Gamma_8, \tag{3.4}$$

$$\Gamma_4 \otimes \Gamma_6 = \Gamma_6 \oplus \Gamma_8, \quad \Gamma_5 \otimes \Gamma_6 = \Gamma_7 \oplus \Gamma_8, \quad \Gamma_6 \otimes \Gamma_6 = \Gamma_1 \oplus \Gamma_4. \tag{3.5}$$

Γ_6 and Γ_7 correspond to $j = \frac{1}{2}$ and identify conduction and the spin-orbit split-off bands, respectively. Γ_8 corresponds to the $j = \frac{3}{2}$ valence band. However, we have to be careful when using the quantum number j to identify discrete representations. This makes sense only for certain small numbers j where there is a clear correspondence, and in general, the Γ notation should be used. Let us elaborate on this correspondence.

Since $T_d \subset O(3)$, we can easily find representations of T_d by evaluating special rotations and reflections of $O(3)$. However, the representations that we find by starting from the (irreducible) $\mathcal{D}^{(j)\pm}$ of $O(3)$, or its cover $SU(2) \times C_i$, may become reducible, when viewed as representations of the discrete group T_d. The decomposition of the $O(3)$ representations is known as compatibility relations, and is shown in Table 3.1. Since irreducible representations are associated with degenerate energy levels, this means that we can expect the lowering of the symmetry (here due to the crystal lattice, reducing the amount of rotations), to appear as splitting of energy levels.

[2] Sometimes a different notation is found for the irreducible representations of T_d, with $\Gamma_{12} \equiv \Gamma_3$, $\Gamma_{15} \equiv \Gamma_4$ and $\Gamma_{25} \equiv \Gamma_5$.

[3] One should not confuse the Γ symbols used for the irreducible representations, and the basis of Γ matrices appearing in the Clifford algebra of Section 7.1, where the method of invariants is applied to a QW system.

$\mathcal{D}^{(0)+}$	Γ_1	$\mathcal{D}^{(0)-}$	Γ_2
$\mathcal{D}^{(1)+}$	Γ_4	$\mathcal{D}^{(1)-}$	Γ_5
$\mathcal{D}^{(2)+}$	$\Gamma_3 \oplus \Gamma_5$	$\mathcal{D}^{(2)-}$	$\Gamma_3 \oplus \Gamma_4$
$\mathcal{D}^{(3)+}$	$\Gamma_2 \oplus \Gamma_4 \oplus \Gamma_5$	$\mathcal{D}^{(3)-}$	$\Gamma_1 \oplus \Gamma_4 \oplus \Gamma_5$
$\mathcal{D}^{(1/2)+}$	Γ_6	$\mathcal{D}^{(1/2)-}$	Γ_7
$\mathcal{D}^{(3/2)+}$	Γ_8	$\mathcal{D}^{(3/2)-}$	Γ_8
$\mathcal{D}^{(5/2)+}$	$\Gamma_7 \oplus \Gamma_8$	$\mathcal{D}^{(5/2)-}$	$\Gamma_6 \oplus \Gamma_8$

Table 3.1: Compatibility relations for T_d with the full rotation group $O(3) = SO(3) \times C_i$ (integer j) and its cover $SU(2) \times C_i$ (for half-integer j). Taken from [Dre+08].

The half-integer representations are only of interest in presence of spin orbit coupling, and decompose into the double group representations $\Gamma_{6/7/8}$. A spin described by a vector in the Hilbert space of half-integer j, will feel the reduction of symmetry due to the lattice only via a spin-orbit coupling term, if present. Rotations in the physical space $\mathbb{R}^3$ of the lattice must be elements of $SO(3)$, with integer represention l, so the upper half of Table 3.1 should be sufficient to understand the physics, even in the spin-orbit coupled case. We see that in T_d, only representations with $l = 2$ or greater become reducible - but we will only consider material properties where s- or p-orbitals are relevant. Thus, we will need only Γ_1 (for s-orbitals) and Γ_5 of T_d or Γ_{15} of O_h (for p-orbitals)[4]. Table 3.1 shows that for these, the irreducible representations of the discrete rotations are identical to the irreducible representations of the full rotation group. This matches the observation that an effective theory can have higher symmetry than the underlying system. An effective theory for a zinc blende system, which is derived only with tensors of $l = 0$ and $l = 1$ (and does not include outer products of those), will have full rotational symmetry. Note that the compatibility table for O_h changes in a non-trivial way, as compared to Table 3.1 [Dre+08]. These tables may be derived using the character tables of a group.

According to the left side of (3.1), the representation of $SU(2)$ appears twice. Therefore, the left hand side transforms according to the tensor product of representations $\mathcal{D}(g) \otimes \mathcal{D}^*(g)$. We can see this explicitly by regarding the matrix H as vector $\mathbf{H}$ with components H_{ij}, and rewriting the transformed Hamiltonian in the form

$$(\mathcal{D}H\mathcal{D}^\dagger)_{il} = \sum_{j,k} \mathcal{D}_{ij} H_{jk} \mathcal{D}^*_{lk} = \sum_{j,k} (\mathcal{D} \otimes \mathcal{D}^*)_{il,jk} H_{jk} = (\mathcal{D} \otimes \mathcal{D}^* \cdot \mathbf{H})_{il}. \tag{3.6}$$

This is useful for shifting a unitary transformation of $\mathcal{D} \otimes \mathcal{D}^*$ to $\mathbf{H}$. We are particularly interested in the unitary transformation that block-diagonalizes the tensor product, resulting in the decomposition into irreducible representations. Usually, the decomposition into irreducible blocks is written like $\mathcal{D}^{(1)} \otimes \mathcal{D}^{(1/2)*} = \mathcal{D}^{(1)} \oplus \mathcal{D}^{(0)}$ etc. From the addition of angular momenta j_1 and j_2, it is well known that all representations $j_1 + j_2, j_1 + j_2 - 1, ..., |j_1 - j_2|$ appear in the product representation. Since it appears here, we want to say a word about the complex conjugate representation [Pik+74]. If the characters of an irreducible representation $\mathcal{D}$ are all real, $\mathcal{D}$ and $\mathcal{D}^*$ are equivalent, i.e. may be transformed into each other by a unitary transformation. The character of a group element g in a particular representation is defined as

$$\chi(\mathcal{D}(g)) = \mathrm{Tr}(\mathcal{D}(g)) \tag{3.7}$$

[4] The p-orbitals transform like the vector (x, y, z). While this corresponds to Γ_5 in T_d, it is Γ_4^- in O_h, and one often finds the notation $\Gamma_{15} \equiv \Gamma_4^-$.

and therefore, it is the same for all equivalent representations. In particular, all characters of the $SU(2)$ representations may be obtained easily from the generator $J_z^{(j)}$, since two rotation vectors of the same length but pointing in different directions are related by a unitary transformation. One finds for a group element g associated with a rotation by angle α,

$$\chi(\mathcal{D}^{(j)}(g)) = \operatorname{Tr} e^{i\alpha J_z^{(j)}} = \sum_{m=-j}^{j} e^{i\alpha m} = \frac{\sin((j+\frac{1}{2})\alpha)}{\sin(\frac{\alpha}{2})}, \tag{3.8}$$

so indeed the characters are all real. Explicitly, we find the $\mathcal{D}^{(j)}$ and $\mathcal{D}^{(j)*}$ are related by

$$e^{i\pi J_y^{(j)}} \mathcal{D}^{(j)} e^{-i\pi J_y^{(j)}} = \mathcal{D}^{(j)*}. \tag{3.9}$$

This means that we do not need to worry much, whether we are searching for the decomposition into irreducible representations of the product $\mathcal{D}^{(j_1)} \otimes \mathcal{D}^{(j_2)}$, $\mathcal{D}^{(j_1)*} \otimes \mathcal{D}^{(j_2)}$ or $\mathcal{D}^{(j_1)} \otimes \mathcal{D}^{(j_2)*}$, the result is the same up to a global (i.e. independent of g) unitary transformation. Of course, once we have adopted a basis convention and we are looking for an explicit set of matrices for a representation, the complex conjugation is important.

Let's reconsider Eq. (3.6). We assume that $\mathcal{D}$ has dimension $n \times n$, and further assume that the unitary g-independent $n^2 \times n^2$-matrix V transforms $V^\dagger(\mathcal{D} \otimes \mathcal{D}^*)V$ into irreducible blocks, each block equal to some $\mathcal{D}^{(q)}$. Here and in the following, we use the dot $\cdot$ to denote the matrix-vector product in the n^2-dimensional space, i.e. $\left[V^\dagger \cdot \mathbf{H}\right]_{ab} = \sum_{c=1}^{n}\sum_{d=1}^{n} V^\dagger_{ab,cd} H_{cd}$. We use the block structure of $V^\dagger(\mathcal{D} \otimes \mathcal{D}^*)V$ to define new basis vectors $\hat{e}_{qm}$ and find coefficients $H_m^{(q)}$ by requiring

$$V^\dagger \cdot \mathbf{H} = \sum_q \sum_{m=-q}^{m} \hat{e}_{qm} H_m^{(q)}. \tag{3.10}$$

In the original basis, the basis vectors look like $n \times n$-matrices $\hat{E}_m^{(q)} = V \cdot \hat{e}_{qm}$. We will see that these matrices transform under rotations like the components of a spherical q-tensor. The expansion of H in the spherical tensor basis looks like $H = \sum_{q,m} \hat{E}_m^{(q)} H_m^{(q)}$. We have

$$\mathcal{D}^\dagger \hat{E}_m^{(q)} \mathcal{D} = (\mathcal{D} \otimes \mathcal{D}^*)^\dagger \cdot V \cdot \hat{e}_{qm} = V \cdot \mathcal{D}^{(q)\dagger} \cdot \hat{e}_{qm} = V \cdot \sum_{m'} \mathcal{D}^{(q)*}_{mm'} \hat{e}_{qm'} = \sum_{m'} \mathcal{D}^{(q)*}_{mm'} \hat{E}_{m'}^{(q)} \tag{3.11}$$

Note the different order of $\mathcal{D}$, $\mathcal{D}^\dagger$ on the left hand side, compared to (3.1). The difference appears because the transformation has been applied to basis vectors instead of vectors of coefficients. Eq. (3.11) matches the definition of a spherical rank-q tensor [Sak04]. q will be always integer, even if half-integer representations appear in $\mathcal{D}$. Thus it is possible to form scalar products with vectors of the physical space $\mathbb{R}^3$ to obtain invariant expressions, even when the spinor transforms like half-integer j.

As simplest example, consider the Zeeman term for a spin-$\frac{1}{2}$ in a magnetic field. The scalar product of Cartesian vectors can be written as scalar product of spherical tensors, $\boldsymbol{\sigma} \cdot \mathbf{B} = \boldsymbol{\sigma}^\dagger U^\dagger U \mathbf{B} = \sigma_{-1}^\dagger B_{-1} + \sigma_0^\dagger B_0 + \sigma_1^\dagger B_1$, where we have introduced the unitary matrix U that transforms a Cartesian 3-vector into a spherical rank-1 tensor,

$$\begin{pmatrix} B_{-1} \\ B_0 \\ B_{+1} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} = U \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix}, \tag{3.12}$$

and likewise for $\boldsymbol{\sigma}$. The spherical tensor components correspond to the commonly used linear combinations $\sigma_\pm = (\sigma_x \pm i\sigma_y)/2$, but the prefactors are different because we want U to be unitary. To avoid confusion, we use the notation $\sigma_{\{-1,0,1\}}$ instead of $\sigma_{\{-,z,+\}}$. So when expanding the Zeeman term in the spherical basis, $B_{\{-1,0,1\}}$ play the role of the coefficients $H^{(1)}_{\{-1,0,1\}}$ and the $\sigma_{\{-1,0,1\}}$ play the role of the spherical rank-1 basis matrices $\hat{E}^{(1)}_{\{-1,0,1\}}$. While the basis matrices are fully determined by the diagonalization matrix V, which is essentially given by the Clebsch-Gordan coefficients [Sak04], the expansion coefficients are arbitrary parameters, here given by the components of the magnetic field. Both transform according to the same irreducible representation $\mathcal{D}^{(1)}$, which is a general principle. We can only form invariant scalar products if the tensors transform equally. Another observation that is true in general, is that the spherical components are complex-valued and usually not Hermitian, while the Cartesian tensor components are real coefficients or Hermitian basis matrices. Therefore, one usually prefers to work with the Cartesian expressions. Also for rank $l = 2$ and higher, one can find the Cartesian components in a similar manner, by forming linear combinations of spherical components $-m$ and m. The spherical tensor components behave analogue to the spherical harmonics $Y^{(l)}_m$, and the linear independent functions $Y^{(l)}_m$ and $Y^{(l)}_{-m}$ are always related by complex conjugation.

To relate Cartesian 3-vector rotation operators to their spherical representation, we have the very useful relation

$$URU^\dagger = \mathcal{D}^{(1)*}(R), \tag{3.13}$$

where R represents a real orthogonal 3×3 rotation matrix and the abstract group element at the same time. (3.13) can be shown by using $R = e^{\boldsymbol{\omega}\cdot\mathbf{R}}$ with the antisymmetric generators $(R_j)_{kl} = \epsilon_{jkl}$, and analysing their relation to the generators $J^{(1)}_{x/y/z}$. Using the relation (3.12) between Cartesian and spherical tensors, (3.11) and (3.13), one finds

$$\mathcal{D}^{(1/2)}(R)\,\boldsymbol{\sigma}\,\mathcal{D}^{(1/2)\dagger}(R) = U^\dagger \mathcal{D}^{(1)*}(R^T) U \boldsymbol{\sigma} = R^T \boldsymbol{\sigma}, \tag{3.14}$$

and more generally, for any j,

$$\mathcal{D}^{(j)}(R)\,\mathbf{J}^{(j)}\,\mathcal{D}^{(j)\dagger}(R) = R^T \mathbf{J}^{(j)}. \tag{3.15}$$

The same transformation rule applies to $\mathbf{B}$, if we assume that $\mathcal{D}(R)$ also acts on $\mathbf{B}$. On the other hand, if $\mathbf{B}$ is seen as parameter, we may write $H(\mathbf{B}) = \boldsymbol{\sigma} \cdot \mathbf{B}$ and then we find $\mathcal{D}(R)H(R^T\mathbf{B})\mathcal{D}^\dagger(R) = H(\mathbf{B})$, as is required by the invariance condition (3.1).

A general invariant Hamiltonian can be written by forming all possible scalar products of tensors that transform equally under rotations, including all possible tensors up to some order in $\mathbf{k}$ and other parameters. This result is also true when the irreducible representations do not match exactly the $SU(2)$ representations because those are reducible. Therefore, we switch to a different notation, emphasizing the use of representations of discrete point groups. The dimension of a representation Γ_κ is denoted as L_κ instead of $2j + 1$. Introducing additional indices λ, μ to distinguish terms that transform according to the same representation Γ_κ, but have different physical content, we may write any possible invariant as [Win05]

$$I_{\kappa\lambda\mu} = \sum_{l=1}^{L_\kappa} X_l^{(\kappa,\lambda)} \mathcal{K}_l^{(\kappa,\mu)*} \tag{3.16}$$

where usually, the $\mathcal{K}_l^{(\kappa,\mu)}$ are regarded as spherical tensor components of parameter vectors, and the $X_l^{(\kappa,\lambda)}$ are regarded as the set of basis matrices implementing the l-components of a spherical tensor. However, one does not have to distinguish strictly between parameters and basis matrices, e.g. sometimes one could regard the $\mathcal{K}_l^{(\kappa,\mu)}$ as matrices as well. One only has to ensure that both parts transform under rotation, and it is less important whether the rotation is implemented in the Hilbert space or parameter space. Further, if the band basis of a Hamiltonian contains more than one representation, the method can also be generalized to diagonal and off-diagonal blocks of H. In this case, one labels the blocks according to the representation $\Gamma_\alpha \otimes \Gamma_\beta^*$ by which they transform. The index κ runs over the representations contained in the tensor product. The general invariant block of the Hamiltonian can be constructed by assigning an arbitrary real prefactor $a_{\kappa\lambda\mu}^{\alpha\beta}$ to all possible invariants, giving [Win05]

$$H_{\alpha\beta}(\mathcal{K}) = \sum_{\kappa,\lambda,\mu} a_{\kappa\lambda\mu}^{\alpha\beta} \sum_{l=1}^{L_\kappa} X_l^{(\kappa,\lambda)} \mathcal{K}_l^{(\kappa,\mu)*}. \tag{3.17}$$

Up to now, we have limited our discussion to transformations contained in $SU(2)$. But the terms that may appear in (3.17) are further restricted by transformation properties of tensor components under parity and time reversal, if these are symmetries of the system. If parity or time reversal symmetry are broken only because of externally applied electric or magnetic fields, we can include the transformation of the fields in the symmetry operations and still obtain useful restrictions of the allowed terms. Even if the point group of the crystal lacks inversion symmetry, like e.g. T_d, it is useful to classify terms by their parity. To this end, it is common to use basis states of a corresponding inversion symmetric crystal, and to consider terms breaking the inversion symmetry as perturbation. In particular, adding the inversion element to T_d gives the group O_h which is the point group of the diamond structure. Since we work with a basis of parity eigenstates, the irreducible representation corresponding to each basis set should be assigned a parity index, like e.g. Γ_α^+ or Γ_β^-. The parity of the contained representations Γ_κ^- is simply determined by the product of the parities. The parity is not determined by the original $SU(2)$ representation - the same representation may exist for both parity values ± 1. In a parity-even system, we may only combine representations (κ,λ) and (κ,μ) of equal parity in (3.17). In zinc blende systems, there are additional parity-odd terms called bulk inversion asymmetry (BIA) terms.

Also, the time reversal operator $\mathcal{T}$ gives further restrictions to the terms allowed in (3.16) and (3.17). However, this applies only to the diagonal blocks $H_{\alpha\alpha}$. One can classify all the basis matrices $X_m^{(\kappa,\lambda)}$ into even or odd transformation under time reversal, i.e. $\mathcal{T} X_m^{(\kappa,\lambda)} \mathcal{T}^{-1} = \pm(-1)^m X_{-m}^{(\kappa,\lambda)}$ (applying time reversal to spherical tensors switches components $-m \leftrightarrow m$, but this is unimportant when forming a scalar product). Likewise, the spherical tensors $\mathcal{K}^{(\kappa,\mu)*}$ are classified according to their even/odd behaviour under time reversal. Then, only even products are allowed. The off-diagonal blocks $H_{\alpha\beta}$, $\alpha \neq \beta$, are not restricted by time reversal, which becomes clear from the following argument. The time reversal operator $\mathcal{T}$ maps basis states $|j,m\rangle$ onto $\pm|j,-m\rangle$ of the same representation. However, the sign of $\mathcal{T}$ is arbitrary and only fixed by convention. It does not matter whether we consider $\mathcal{T}$ or $-\mathcal{T}$ as time reversal operator, all that is important is the sign of $\mathcal{T}^2$, which should be -1 for j half-integer and $+1$ for j integer. Since different representations are not linked by $\mathcal{T}$, we may use different sign conventions for e.g. the Γ_6 conductance and Γ_8 valence bands, increasing the arbitrariness of the combined time reversal operator. This means that we cannot assign an even/odd behaviour under time reversal

to a $\Gamma_6 \otimes \Gamma_8^*$ block, and $\mathcal{T}$ gives no restriction to the allowed elements of the off-diagonal blocks.

3.1 Effective quadrupole Hamiltonian by symmetry

We will explicitly elaborate the example of rank-2 basis matrices transforming as a quadrupole, since this will be relevant for the Hamiltonian of a spin-$\frac{3}{2}$ in an electrical quadrupole field, a subject of Chapter 8. Let's start with an arbitrary 3×3 matrix A, which may be complex valued and does not even have to be Hermitian. The matrix is assumed to transform like a (reducible) Cartesian rank-2 tensor under rotations, i.e. $A \to RAR^T$ where R is the rotation matrix. If an electrical quadrupole potential is implemented as $\mathbf{r}^T A \mathbf{r}$, this is the needed transformation property. The orthogonal decomposition in parts transforming under rotations according to different representations is [Sak04]

$$A = A^{(0)} \oplus A^{(1)} \oplus A^{(2)} \equiv \frac{\mathrm{Tr}A}{3}\mathbb{1} \oplus \frac{1}{2}(A - A^T) \oplus \left(\frac{1}{2}(A + A^T) - \frac{\mathrm{Tr}A}{3}\mathbb{1}\right) \tag{3.18}$$

where the first part, the trace times identity, transforms like $l = 0$, the second part, which is antisymmetric and can be seen as 3-component vector, transforms like $l = 1$, and the third part is symmetric and traceless and transforms like $l = 2$, having 5 independent components. The direct sum sign $\oplus$ indicates orthogonality under the scalar product of matrices

$$\mathrm{Tr}(A^{(i)}A^{(j)}) = \mathbb{1}\delta_{ij}. \tag{3.19}$$

as can be shown by using the cyclic invariance under the trace.

The number of parameters in the decomposition (3.18) add up as $3 \times 3 = 1 + 3 + 5$, just like the correct decompositon into spherical tensor components. But this is not yet a proof of irreducibility. To make clear that the $A^{(l)}$ are indeed irreducible parts, we will now show explicitly, how to construct an irreducible $l = 2$ basis by symmetry, and then see that the basis is indeed traceless and symmetric.

To do so, there are two ways. Firstly, we can use the defining invariance condition (3.11) for a spherical tensor, with $\mathcal{D}^{(1)}$ on the left and $\mathcal{D}^{(2)}$ on the right hand side, where $\mathcal{D}^{(j)}(R) = e^{i\boldsymbol{\omega}\mathbf{J}^{(j)}}$. For arbitrary infinitesimal rotations, this gives

$$[J_i^{(1)}, \bar{Q}_j] = \sum_{m=-j}^{j} (J_i^{(2)*})_{jm}\bar{Q}_m, \quad i \in \{1,2,3\},\ j \in \{-2,...,2\} \tag{3.20}$$

which is a system of $3 \cdot 5 \cdot 9 = 135$ equations for the $5 \cdot 9$ components of the spherical tensor components $\bar{Q}_m$, which should transform according to $l = 2$. One extra equation is needed for a normalization condition of the $\bar{Q}_m$, and this will also exclude the trivial solution $\bar{Q}_m = 0$. The system can be readily solved with MATHEMATICA, giving a solution set

$$\bar{Q}_{-2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \sqrt{\frac{3}{5}} & 0 & 0 \end{pmatrix}, \quad \bar{Q}_{-1} = \begin{pmatrix} 0 & 0 & 0 \\ -\sqrt{\frac{3}{10}} & 0 & 0 \\ 0 & \sqrt{\frac{3}{10}} & 0 \end{pmatrix}, \quad \bar{Q}_0 = \begin{pmatrix} \frac{1}{\sqrt{10}} & 0 & 0 \\ 0 & -\sqrt{\frac{2}{5}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{10}} \end{pmatrix},$$
$$\bar{Q}_1 = \begin{pmatrix} 0 & \sqrt{\frac{3}{10}} & 0 \\ 0 & 0 & -\sqrt{\frac{3}{10}} \\ 0 & 0 & 0 \end{pmatrix}, \quad \bar{Q}_2 = \begin{pmatrix} 0 & 0 & \sqrt{\frac{3}{5}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.21}$$

Note that these are not yet the symmetric basis matrices that we are looking for. We still need to apply some basis transformations. But before we do so, let's show a different method to construct the same set of basis matrices, which is based on the Wigner-Eckart theorem [Sak04]. The theorem relates matrix elements $\langle \alpha, jm|T_k^{(q)}|\alpha', j'm'\rangle$ to one another. Here, $T_k^{(q)}$ is the component k of a spherical rank-q tensor. α and α' are some general quantum numbers, and j, m and j', m' are angular momentum quantum numbers,

$$\langle \alpha, jm|T_k^{(q)}|\alpha', j'm'\rangle = \langle j'q, m'k|j'q, jm\rangle \frac{1}{\sqrt{2j'+1}} \langle \alpha, j||T^{(q)}||\alpha', j'\rangle. \tag{3.22}$$

The theorem states that the matrix element falls into a factor depending on the magnetic quantum numbers m, m', k, which essentially carry information about vector directions, but this factor does not depend on the operator $T^{(q)}$ or α and α', and a second factor not depending on m, m', k. The first factor is given by the Clebsch-Gordan coefficients $\langle j'q, m'k|j'q, jm\rangle$. We may imagine adding the angular momenta of the tensor and the rhs state, and projecting the result onto the angular momentum of the lhs state. The second factor contains a matrix element $\langle \alpha, j||T^{(q)}||\alpha', j'\rangle$, which needs to be determined only once, to obtain the full matrix for different m, m', k.

Applying this to our case, the 3×3 quadrupole basis matrices should not depend on some α, α', and all we need to do, is to evaluate the Clebsch-Gordan coefficients for $j = j' = 1$ and $q = 2$. The basis matrices $\bar{Q}_k$ for $k \in \{-2, ..., 2\}$ are written in the basis of $m, m' \in \{-1, 0, 1\}$.

Having shown two ways to obtain the spherical tensor basis $\{\bar{Q}_m\}_m$, we use them to construct a Hermitian basis that transforms as Cartesian rank-2 tensor. Note, the construction here ($l = 2$) is more complicated than in the $l = 1$ case, of constructing the Pauli matrix basis. There, taking Hermitian linear combinations of σ_{+1} and σ_{-1} is all that is required to obtain a matrix basis which transforms like a Cartesian vector. But here, we first want to transform the matrix basis of individual components to the Cartesian basis, because the decomposition (3.18) is only correct for the Cartesian basis. The matrices $\{\bar{Q}_m\}_m$ are all real, but this is merely due to a nice phase convention of the Clebsch-Gordan coefficients. More importantly, they have the property $\bar{Q}_m^\dagger = \bar{Q}_m^T = (-1)^m \bar{Q}_{-m}$, similar to the relation $Y_m^{l*} = (-1)^m Y_{-m}^l$ for spherical harmonics. They are not symmetric, and we would like to avoid explicit symmetrisation. Let us consider the basis of

$$\tilde{Q}_m = U^\dagger \bar{Q}_m U, \tag{3.23}$$

which are written in the Cartesian vector basis (x, y, z). They are are no longer real, but are symmetric (can be checked by evaluation), and this proves that the decomposition (3.18) splits up into irreducible Cartesian components. Therefore,

$$\tilde{Q}_m^\dagger = \tilde{Q}_m^* = (-1)^m \tilde{Q}_{-m}. \tag{3.24}$$

We use this to find the Cartesian quadrupole (i.e. $q = 2$) basis matrices, written in the Cartesian basis. They are real, symmetric, traceless and form an orthogonal set with the scalar product (3.19). However, they still contain information about the direction of the coordinate system used for the spherical decomposition.

$$Q_0 = -\frac{\sqrt{10}}{3}\tilde{Q}_0 = \frac{1}{3}\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad Q_1 = -\frac{\sqrt{5}}{3}(\tilde{Q}_{-1} - \tilde{Q}_1) = \frac{1}{\sqrt{3}}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$Q_2 = \frac{\sqrt{5}}{3i}(\tilde{Q}_{-1} + \tilde{Q}_1) = \frac{1}{\sqrt{3}}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad Q_3 = -\frac{\sqrt{5}}{3}(\tilde{Q}_{-2} + \tilde{Q}_2) = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$Q_4 = \frac{\sqrt{5}}{3i}(\tilde{Q}_{-2} - \tilde{Q}_2) = \frac{1}{\sqrt{3}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.25}$$

Being all real, these quadrupole basis matrices are even under time reversal, and they are also even under parity.

We are interested in the simplest possible Hamiltonian that allows to control (by external parameters) the spin of an electron or some fermionic quasiparticle, without breaking time reversal invariance (TRI) [Avr+88; Bud+12a]. Considering that we will have Kramers' degeneracy and we need an energy splitting to have some sort of control, the smallest system that could implement the control, is a $j = \frac{3}{2}$ quadruplet of states, that will be subject to some controlled perturbation. We will now argue that the only possible perturbation Hamiltonian is given by [Avr+89]

$$H(Q) = \sum_{i,j=x,y,z} J_i^{(3/2)} Q_{ij} J_j^{(3/2)} \tag{3.26}$$

where Q is a quadrupole matrix, i.e. a traceless symmetric real matrix given by $Q = \sum_{i=0}^{4} q_i Q_i$ with real coefficients q_i.

The standard way to apply the theory of invariants will be to consider the product $\mathcal{D}^{(3/2)} \otimes \mathcal{D}^{(3/2)*} = \mathcal{D}^{(0)} \oplus \mathcal{D}^{(1)} \oplus \mathcal{D}^{(2)} \oplus \mathcal{D}^{(3)}$, then construct basis matrices for the contained representations by the Wigner-Eckart theorem, and analyse their transformation under time reversal and parity (the parity is positive for all of them because we are considering a diagonal block $H_{\alpha\alpha}$). Then, scalar products with matching tensors should be formed. But here, only time reversal even tensors should be considered, since the resulting Hamiltonian should not contain an external magnetic field that breaks TRI. We note that $\mathcal{D}^{(0)}$ gives only an uninteresting overall energy shift. The basis matrices for $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(3)}$ are found to be odd under time reversal, and thus excluded. Finally, one can write the scalar products in terms of Cartesian components, which are Hermitian. The 5 basis matrices corresponding to $\mathcal{D}^{(2)}$ turn out to be identical to the 4×4-matrices $\sum_{i,j=x,y,z} J_i^{(3/2)} \tilde{Q}_m J_j^{(3/2)}$, $m \in \{-2, ..., 2\}$.

However, we find that it is more elegant to start with (3.26) as an ansatz and then, use the transformation property (3.15) to shift the rotation to the matrix Q, finding $\mathcal{D}(R)H(Q)\mathcal{D}^\dagger(R) = \mathbf{J}^T RQR^T \mathbf{J} = H(RQR^T)$. Then, all possible Cartesian 3×3 matrices Q can be analysed, as has been done above. Using an antisymmetric matrix Q and the commutation relation of the J_i, one can form the Zeeman term $\mathbf{B} \cdot \mathbf{J}$, but this is not of interest here. The term $\mathbf{E} \cdot \mathbf{J}$ is not allowed because it transforms odd under time reversal. However, the square $|\mathbf{E} \cdot \mathbf{J}|^2$ is an allowed term. It can be rewritten in terms of the superposition of a quadrupole Hamiltonian with $Q = \mathbf{E}\,\mathbf{E}^T - \frac{\mathbf{E}^2}{3}\mathbb{1}$ and a constant overall energy shift of $\frac{15}{4}\frac{\mathbf{E}^2}{3} = \frac{5}{4}\mathbf{E}^2$. So in this sense, the term transforms as quadrupole, even if it has only the three parameters of the vector $\mathbf{E}$.

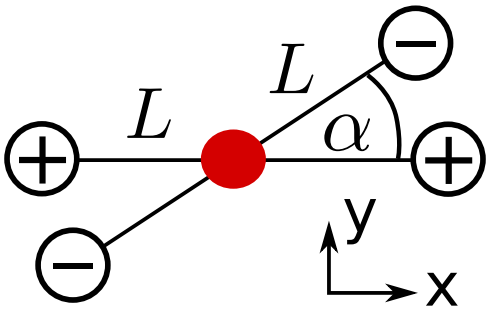


Figure 3.1: A charge configuration in the $x-y$ plane which generates a quadrupole field at the center, assuming a near field approximation where $r \ll L$. Here r is the radius of the interesting area occupied e.g. by a quantum dot, shown in red. The angle α and the field strength are quadrupole parameters which are however invariant under rotations. The other three quadrupole parameters may be seen as Euler angles describing the orientation of the plane occupied by the charges, relative to the embedding coordinate system.

This term corresponds to the quadratic Stark effect. Although the linear Stark effect is not allowed in a Γ_8 band if the system has $SO(3)$ or $O_h = T_d \times C_i$ symmetry, it is allowed in T_d symmetry [Bir+63]. Before we discuss the Stark effects in T_d symmetry, let's look a bit closer at the quadrupole tensors Q and how they parametrize the field.

A quadrupole configuration of charges at distance L from the origin gives rise to an electrostatic potential of the form $\mathbf{r}^T Q \mathbf{r}$ at the origin in near field approximation, for which we must assume $r \ll L$, compare Fig. 3.1. In order to understand the effect of rotations on the quadrupole, we may think of the five-dimensional space of quadrupoles $Q = \sum_{i=0}^{4} x_i Q_i$ being parametrized by the quadrupole strength $\sqrt{\sum_i x_i^2}$, the distortion angle α shown in Fig. 3.1, and three Euler angles parametrizing the orientation of the plane of the four charges (or more generally, the symmetry plane of the charge configuration). Only the Euler angles change under rotations.

3.2 Relation to Stark effect Hamiltonians

Let us consider a system with T_d symmetry and derive the form of the linear and quadratic Stark Hamiltonian from the ansatz (3.26). For this, we consider the specific charge configurations of the Q_i basis matrices to find appropriate $\Gamma_8 \times \Gamma_8^*$ basis matrices which transform like the irreducible representations of T_d. The vector space spanned by Q_0 and Q_3 conforms to quadrupole configurations, where the midpoints of the faces of a cube are charged, e.g. charges q at $(\pm 1, 0, 0)$ and charges $-q$ at $(0, \pm 1, 0)$. Evaluating $\mathbf{J}^T Q \mathbf{J}$, we obtain the basis $(\frac{1}{\sqrt{3}}(2J_z^2 - J_x^2 - J_y^2), J_x^2 - J_y^2)$, which transforms like Γ_3 of T_d [Win05]. On the other hand, the basis (Q_2, Q_1, Q_4) corresponds to charge configurations which are rotated by $\pi/4$ in the xy, yz, zx planes, e.g. charges q at $(\pm 1, \pm 1, 0)$ and $-q$ at $(\pm 1, \mp 1, 0)$. The vector space of possible Hamiltonians $\mathbf{J}^T Q \mathbf{J}$ is spanned by the basis $(\{J_y, J_z\}, \{J_z, J_x\}, \{J_x, J_y\})$, which transforms like Γ_5 of T_d. Distorted quadrupole configurations, where $\alpha \neq \frac{\pi}{2}$ in Figure 3.1, can be obtained by superpositions of the Γ_3 and Γ_5 quadrupole bases.

In order to find the form of the linear Stark effect Hamiltonian, we need tensors linear in $\mathbf{E} = (E_x, E_y, E_z)$. In T_d, the only one is $\mathbf{E}$ itself, and it transforms like Γ_5. The only set of basis matri-

ces which is also even under time reversal and transforms like Γ_5, is $(\{J_y, J_z\}, \{J_z, J_x\}, \{J_x, J_y\})$ [Win05]. Thus, the linear Stark Hamiltonian for the ground state of an acceptor centre in a T_d crystal is given by a single term [Bir+63],

$$H_E = \chi(E_x\{J_y, J_z\} + E_y\{J_z, J_x\} + E_z\{J_x, J_y\}). \tag{3.27}$$

It is parity-odd, and therefore forbidden in the symmetry group $O_h = T_d \times C_i$, which additionally contains the spatial inversion element. Note also, that full rotational symmetry $SO(3)$ would already forbid this term, regardless of inversion symmetry. For the Stark effect, one should regard $\mathbf{E}$ as an externally applied field, and χ is a measure of polarization. The field gives rise to an energy splitting of $2\frac{\sqrt{3}}{\chi}|E|$. But if $\mathbf{E}$ in (3.27) refers to the internal crystal field, then H_E becomes a bulk inversion asymmetry (BIA) Hamiltonian.

Likewise, the quadratic Stark effect is found by taking the quadratic in $\mathbf{E}$ tensors E^2 (of Γ_1), $(\frac{1}{\sqrt{3}}(2E_z^2 - E_x^2 - E_y^2), E_x^2 - E_y^2)$ (of Γ_3) and (E_yE_z, E_zE_x, E_xE_y) (of Γ_5), and combining them with the corresponding basis matrices. In T_d or O_h, the quadratic Stark Hamiltonian for the ground state of the acceptor centre looks like [Bir+63]

$$\begin{aligned} H_{E^2} =& \epsilon E^2 I + \beta\,(E_x^2 J_x^2 + E_y^2 J_y^2 + E_z^2 J_z^2 - \frac{5}{4} I E^2) \\ &+ \delta(E_yE_z\{J_y, J_z\} + E_zE_x\{J_z, J_x\} + E_zE_x\{J_z, J_y\}) \end{aligned} \tag{3.28}$$

where the first term is a trivial shift, and the second (third) term can be expressed in terms of a Γ_3 (Γ_5) quadrupole Hamiltonian. In full rotational symmetry, $\beta = \delta$ is enforced. Neglecting the uninteresting energy shift, we obtain the form $H_{E^2} = \beta\,\mathbf{J}^T(\mathbf{E}\mathbf{E}^T - \frac{E^2}{3}\mathbb{1})\mathbf{J}$, recovering our earlier result. Usually, one would expect to see more energy levels in O_h than in a system with full rotational, thus higher, symmetry. But here, this cannot happen, since Kramers' degeneracy must still hold, allowing for only two energy levels.

Chapter 4

Adiabatic theorem and geometric phases in quantum mechanics

Here we present the adiabatic theorem in its most general and useful form. It gives rise to the Berry phase [Ber84], which in the general non-Abelian case is also called Wilczek-Zee phase [Wil+84]. As application, we then discuss the semiclassical transport theory in terms of wave packets. This theory is compared to the Foldy-Wouthysen transformation, the latter being conceptually different. Other applications of the adiabatic theorem of relevance for us (and which are subjects of later chapters), are the generation of a $SU(2)$ holonomy in spin space for a time dependent but always time reversal invariant Hamiltonian, and a formulation of the topological Chern integer as integral of a Berry curvature.

Consider the time dependent Schrödinger equation of a time dependent Hamiltonian $H(\mathbf{R}(t)) = H(t)$ (with $\hbar = 1$),

$$i\frac{\partial}{\partial t}|\psi(t)\rangle = H(\mathbf{R}(t))|\psi(t)\rangle, \tag{4.1}$$

where $\mathbf{R}$ is a parameter vector which is controlled externally in a time dependent way. $\mathbf{R}$ may parametrize, depending on the problem of interest, external electric or magnetic fields, or some coordinates changing slowly compared to the fast electron dynamics, like e.g. the position of the atomic nuclei in a molecule, i.e. coordinates that are treated classically in the Born-Oppenheimer approximation.

The spectrum of H depends parametrically on $\mathbf{R}$, but for the moment we are more interested in the time dependency, assuming a fixed path $\mathbf{R}(t)$. We assume the spectrum to be gapped at all times t, meaning that the instantaneous eigenvalue of interest, $E(t) = E(\mathbf{R}(t))$, is separated from all other eigenvalues by at least Δ. The instantaneous basis of eigenvectors belonging to the possibly degenerate eigenvalue $E(t)$ is $\{|u_n(t)\rangle = |u_n(\mathbf{R}(t))\rangle\}_n$. Actually, the assumption of a gapped spectrum is non-essential, as pointed out both by Born [Bor+28] and Kato [Kat50], the theorem still holds if symmetry protected level crossings exist. The situation becomes different if there is some residual interaction that opens gaps at the level crossings, i.e. generates anti-crossings[1].

[1] From the exactly solvable Landau-Zener problem $H(t) = vt\sigma_z + a\sigma_x$, one can actually see that in such a situation, for a minimal speed $v \gg a^2/\hbar$, the residual interaction $a\sigma_x$ relevant for the anti-crossing can be ignored, so that the adiabatic theorem applies again.

The exact time dependent problem (4.1) is in general difficult to solve. Formally, we may write the solution as

$$|\psi(t)\rangle = U_S(t)|\psi(0)\rangle \equiv \hat{T} \exp\left(-i \int_0^t dt'\, H(\mathbf{R}(t'))\right) |\psi(0)\rangle \tag{4.2}$$

where $\hat{T}$ is the time ordering operator, and we have defined the exact unitary time evolution operator $U_S(t)$.

The adiabatic theorem states that, if the Hamiltonian is very slowly changed with time, then the solution is well approximated by the instantaneous eigenvector, up to a phase. However, the latter is not just the trivial dynamical phase $\exp(-i \int dt' E(t'))$, but also the Berry phase or its generalization. The theorem can already be understood at a very intuitive level, making use of the time-energy uncertainty principle $\Delta E \Delta t > \hbar/2$, which is usually proven by time-dependent perturbation theory. ΔE is the energy difference for allowed excitations and Δt is the time for the Hamiltonian to change. This corresponds to total time T for changing the parameter vector $\mathbf{R}$. So for $T \gg \hbar/\Delta$, excitations to other levels should be suppressed. Like Born and Kato, we also introduce the scale $s = t/T$. The advantage is that we can discuss the limit $T \to \infty$ for fixed s, i.e. for a fixed spectrum. The adiabatic theorem was originally proven by Born and Fock [Bor+28] for the non-degenerate case, and they have already obtained the Berry phase. However, Born and Fock have argued that by a gauge change, i.e. a time-dependent change of the basis $|u(t)\rangle \to e^{i\gamma(t)}|u(t)\rangle$ by a phase factor, the Berry phase may be removed. This is true in most cases, however the important exception as noticed by Berry [Ber84] is when a loop in the parameter $\mathbf{R}(t)$ is performed, so that the starting and ending basis states are identical. In this situation, the Berry phase is gauge invariant and has physical meaning.

We follow the very transparent and general work of Kato [Kat50] and show the most essential steps. The main assumption is that $\frac{dH}{ds} = T\frac{dH}{dt}$ remains finite while $T \to \infty$. Instead of working with basis states, Kato introduces time dependent projectors $P(t) = \sum_n |u_n(t)\rangle\langle u_n(t)|$ projecting on the (degenerate) subspace. This has the important advantage that the formulation is gauge independent from the beginning. In the following, we use the notation $\frac{d}{ds} ='$. First Kato shows that the differential equation

$$X' = iAX \tag{4.3}$$

with $A = -i[P', P]$, can be used to find the adiabatic evolution of a vector $X(t)$, that initially is in the eigenspace. So if $P(0)X(0) = X(0)$, then $P(s)X(s) = X(s)$. Since P (and thus also P') are Hermitian, A is Hermitian. The solution is

$$X(s) = U_A(s)X(0) \equiv \hat{T} \exp\left(i \int_0^s d\tilde{s}\, A(\tilde{s})\right) X(0), \tag{4.4}$$

where $\hat{T}$ is the time-ordering operator. This defines the unitary adiabatic time evolution operator $U_A(s)$, with $U_A(0) = 1$. $U_A(s)$ also obeys the differential equation (4.3).

Now we show that $U_A(s)$ really is the adiabatic evolution. From differentiating $P^2 = P$, we get the important relation

$$PP'P = 0. \tag{4.5}$$

Using this relation one obtains $iPA = -PP'$, $iAP = P'P$ and

$$P' = P'P + PP' = i[A, P]. \tag{4.6}$$

Defining $W(s) = P(s)U_A(s)$, we have

$$W' = (PU_A)' = (P' + iPA)U_A = iA(PU_A) = iAW. \tag{4.7}$$

So $W(s)$ also obeys the differential equation (4.3), and therefore $W(s) = U_A(s)W(0) = U_A(s)P(0)$, so

$$P(s) = U_A(s)P(0)U_A(s)^\dagger. \tag{4.8}$$

This means that indeed, (4.3) gives the adiabatic time evolution.

The next step in Kato's work is to show that for $T \to \infty$ but fixed $s = \frac{t}{T}$, the difference between the adiabatic time evolution and the exact time evolution vanishes like $\frac{1}{T}$,

$$\left(U_S(t) - e^{-i\int_0^t dt' E(t')} U_A(\frac{t}{T})\right) P(0) = \mathcal{O}(T^{-1}). \tag{4.9}$$

The operator A acts on the complete space of eigenstates and projects the eigenspace of $E(t)$ onto its complement (given by the projector $1 - P$) and vice versa, but matrix elements connecting eigenstates of $E(t)$ vanish since $PAP = 0$. Instead of calculating the full U_A we would prefer to know the time evolution only in the eigenspace of $E(t)$. We use the ansatz $X(t) = \sum_i c_i(t)|u_i(t)\rangle$, with coefficients $c_i(t) = c_i(sT)$ which we group into a vector $\mathbf{c}(t)$. From (4.3) we obtain the differential equation for the coefficient vector

$$\frac{d}{dt}\mathbf{c}(t) + iA^B(t)\mathbf{c}(t) = 0 \tag{4.10}$$

with the Hermitian matrix $A^B_{ij}(t) = -i\langle u_i(t)|\frac{d}{dt}|u_j(t)\rangle$. $A^B(t)$ is also called Berry connection, and it can be seen as local (basis-dependent) counterpart to the global (gauge-independent) connection A. The solution vector is $\mathbf{c}(t) = \hat{T}\exp\left[-i\int_0^t dt' A^B(t')\right]\mathbf{c}(0)$. Using $\frac{d}{dt} = \sum_\alpha \frac{dR_\alpha}{dt}\frac{\partial}{\partial R_\alpha}$, we can rewrite the connection $A^B(t)$ in terms of the vector-valued Berry connection $\mathbf{A}^B$ with components A^B_α, defined by

$$A^B_{\alpha,ij} = -i\langle u_i|\frac{\partial}{\partial R_\alpha}|u_j\rangle. \tag{4.11}$$

For a closed path γ in parameter space, $\mathbf{R}(T) = \mathbf{R}(0)$, the operator

$$\hat{T}e^{-i\oint_\gamma \mathrm{d}t'\, A^B(t')} = \hat{T}e^{-i\oint_\gamma \mathrm{d}\mathbf{R}\cdot\mathbf{A}^B} \tag{4.12}$$

is called the holonomy or generalized Berry phase, and the possible holonomy operations form a group (by concatenation of paths, or matrix multiplication). Since in general, the $A^B(t)$ at different times do not commute, the time ordering operator $\hat{T}$ is needed, so (4.12) should only be regarded as formal solution, while an explicit evaluation will be more difficult than a simple integration. However, specializing to the Abelian case $\dim P = 1$, time ordering is not needed, $A^B(t)$ becomes simply a real number, and $e^{-i\int dt' A^B(t')}$ is called Berry phase. Note that for $E(t) \neq 0$, we still need to add the dynamical phase if we want to approximate the solution of the Schrödinger equation, compare Eq. (4.9). The Berry phase is a gauge invariant quantity with possible physical consequences. Let us comment shortly on the physical observability of the holonomy. It is possible to construct interference experiments which detect the (simple) Berry phase [Ber84]. For the generalized Berry phase being gauge dependent, it is less clear if it is

experimentally detectable [Bud12]. While local basis changes along the path γ will cancel out, the holonomy still depends on the choice of basis at the starting and ending point. However, if this choice of basis is fixed by the orientation of the coordinate system in the experiment, the matrix-valued holonomy can be detectable. This is the reason why we can make use of the holonomy as matrix operation, for the spin control by electrical quadrupole fields, a subject of the later Chapter 8.

In order to apply Stokes' theorem of vector calculus, we restrict ourselves to the case of a 3-dimensional (or 2-dimensional) parameter manifold. Considering that the physics remains invariant under re-parametrization of the path in parameter space, it is clear that this is no real restriction. Mathematically, a generalization of Stokes' theorem exists in terms of the outer derivative, which requires the notations of differential forms and the antisymmetric wedge product [Nak03]. Going back to the case $\dim P = 1$, we obtain

$$\oint_\gamma \mathrm{d}t\, A^B(t) = \oint_\gamma \mathrm{d}\mathbf{R} \cdot \mathbf{A}^B = \int_{S(\gamma)} \mathrm{d}\mathbf{S} \cdot \mathbf{F}^B, \tag{4.13}$$

where we have introduced the Berry curvature corresponding to (4.11),

$$\mathbf{F}^B = \nabla \times \mathbf{A}^B = -i\epsilon_{\alpha\beta\delta}\mathbf{e}_\delta \left(\langle \partial_\alpha u | \partial_\beta u\rangle - \langle \partial_\beta u | \partial_\alpha u\rangle\right) = 2\epsilon_{\alpha\beta\delta}\mathbf{e}_\delta \mathrm{Im}\left(\langle \partial_\alpha u | \partial_\beta u\rangle\right) \tag{4.14}$$

with the short notation $|\partial_\alpha u\rangle = \frac{\partial |u\rangle}{\partial R_\alpha}$. $\epsilon_{\alpha\beta\delta}$ is the totally antisymmetric tensor and the $\mathbf{e}_\delta$ are unit vectors, and $S(\gamma)$ is some surface surrounded by the path γ. In contrast to the Berry connection, the Berry curvature is invariant under gauge changes $|u\rangle \to e^{i\phi(\mathbf{R})}|u\rangle$ with some arbitrary real $\phi(\mathbf{R})$, as can be checked easily. Clearly, this is an advantage for practical evaluation of the Berry phase, since a global smooth phase definition for the eigenstates is no longer required.

The mathematical construct for the description of the coefficient vector $\mathbf{c}(\mathbf{R})$ is a fibre bundle [Nak03]. The fibre bundle consists of the parameter manifold as base manifold, with a vector space corresponding to the space of eigenstates of the (possible degenerate) eigenvalue $E(\mathbf{R})$ attached to each point $\mathbf{R}$. This vector space is called the fibre. Fibres which are attached at different points $\mathbf{R}, \mathbf{R}'$ are related by transition functions. These describe the way that fibres at neighbouring points are connected, and the term Berry connection used above indicates that it has to do with the transition functions. For points in parameter space which are infinitesimally close to each other, the transition function becomes a linear map, and is close to identity. Therefore, a transition function connecting points $\mathbf{R}, \mathbf{R}'$ may be generated by a matrix exponential of the integral over linear transition functions, along a path connecting $\mathbf{R}, \mathbf{R}'$. Note that the transition function depends on the path, otherwise the holonomy (4.12) would be trivial. This reveals the geometric interpretation of the Berry connection for the adiabatic transport, and the associated holonomy.

Another simple example of a fibre bundle is a sphere, with the two-dimensional space of tangent vectors attached to each point of the surface. Then, parallel transport of a vector along the surface defines the connection. Since the parallel transport keeps the length of a vector constant, the transition functions are elements of SO(2). Clearly, the defect angle found after parallel transport around a closed loop (and thus the holonomy, which is an element of SO(2)) is proportional to the area enclosed by the loop divided by the radius of the sphere. Since the inverse radius of the sphere is a measure of its curvature, this analogy also motivates the term Berry curvature, as defined by Stokes' theorem in (4.13).

An especially interesting case is when the parameter manifold is compact and has no boundary, like e.g. a sphere or a torus, and $S(\gamma)$ in (4.13) is the complete parameter manifold. Naively,

one would expect the integral (4.13) to be zero. However, this is not true since the Stokes' theorem may be applied only locally, an example will be discussed in Chapter 6.1 in the context of the quantum Hall effect. Instead, the value of the integral becomes quantized and depends only on the topology of the fibre bundle, i.e. on the way it is internally twisted. The integral of the Berry curvature over the complete parameter manifold is known as Chern integral and the topological integer is known as first Chern number. Again, parallel transport and geometric curvature provide an illustrative example - the Gauss-Bonnet theorem states that the integral over the Riemann curvature is given by the Euler characteristic of the surface, which is a topological invariant and may be calculated from a triangulation of the surface (i.e. a mesh of triangles representing the surface), as number of vertices plus number of faces minus number of edges of the mesh. The Gauss-Bonnet states that for a surface S without boundary, $\int_S K = 2\pi\chi$, where K is the Gaussian curvature (if a sphere of radius R can be locally fitted to the surface, $K = 1/R^2$), and χ is the Euler characteristic.

The Chern integer will be discussed in more detail in Section 6, since it explains the quantization of the quantum Hall conductance. The quantum spin Hall effect, which is also a topological effect, requires a more complex theory than the first Chern integer for its explanation. The difficulty is to include the essential restriction that time reversal symmetry requires, into the theory of fibre bundles. However, by means of perturbation theory, the problem may be simplified, allowing also an explanation of the quantum spin Hall effect in terms of first Chern integers.

Note that for the degenerate case, $\dim P > 1$, we cannot use Stokes' theorem to link the generalized Berry phase to an integral over a curvature. However, there still exists a generalization of the Berry curvature, which can be deduced from the parallel transport along the edges of a parallelogram in parameter space, applied to the formalism of fibre bundles. The generalized Berry curvature is given by

$$\mathcal{F}^B_{\mu\nu} = \left(\partial_\mu A^B_\nu - \partial_\nu A^B_\mu\right) + i[A^B_\mu, A^B_\nu] \tag{4.15}$$

where the first term is the immediate generalization of (4.14), while the commutator of connections is new but vanishes in the Abelian case $\dim P = 1$. Under gauge changes (i.e. changes of the local basis of the fibre, given by an $\mathbf{R}$-dependent unitary matrix V), it transforms like $\mathcal{F}^B_{\mu\nu} \to V\mathcal{F}^B_{\mu\nu}V^\dagger$ [Nak03]. We will encounter the generalized Berry curvature below, when applying the adiabatic theorem to derive the semiclassical equations of motion for a wave packet.

4.1 Semiclassical theory of wave packets

We now apply the adiabatic equation of motion (4.10) in order to obtain the semiclassical equations of motion for a wave packet centred at coordinates $\mathbf{x}_c$, $\mathbf{k}_c$, as given in [Sun+99] in the non-degenerate case and in [Cul+05] for the case of a degenerate or quasi-degenerate electronic band. In the semiclassical approximation it is assumed that the wave packet stays within the same band, i.e. inter-band transitions are neglected - matching the regime of validity of the adiabatic theorem. This is a good approximation if the band of interest is energetically well separated from other bands, and if the kinetic energy is low enough that space-dependent perturbations do not give rise to band transitions. Further, by describing the wave packet motion only by classical center coordinates $\mathbf{x}_c$, $\mathbf{k}_c$, one neglects the time-dependent spreading of the wave packet. If one is interested in an effective quantum mechanical theory that also models the spreading of the wave packet, one may apply a re-quantization procedure for the center coordinates. But this is difficult in the general case, as will be discussed below. The

semiclassical method can be seen as an alternative approach to the envelope function method, since it also allows inclusion of slowly varying perturbations of the periodic crystal potential, in order to obtain effective equations of motion. It takes more a geometrical and less a quantum mechanical point of view. For the generalized theory for the (quasi-)degenerate band, the time evolution of the spinor in the degenerate band still obeys an effective Schrödinger equation. Thus, we observe that the approach introduces a separation of time scales for the (fast) spinor motion and the (slow) center coordinates motion. Here, the terms slow and fast should be understood in the sense that the slow motion does not cause band transitions out of the (quasi)degenerate subset for the fast degree of freedom. It is not really necessary for the spinor motion to e.g. carry out a fast precession. The terms slow/fast are intuitive, though, considering that in the Born-Oppenheimer approximation, which treats the motion of nuclei as adiabatic, the motion of the electrons is fast compared to the nuclei.

We assume that the full Hamiltonian contains a locally lattice-periodic crystal potential, but due to some perturbations, a modulation of the strict periodicity may appear on a length scale much larger than the lattice constant. Depending on the problem of interest, the perturbation may be due to external electric fields, magnetic fields included by the vector potential, or a local deformation of the lattice etc. Formally, we may write the Hamiltonian in the form $\hat{H}[\hat{\mathbf{x}}, \hat{\mathbf{k}}, \beta(\hat{\mathbf{x}}, t)]$, where $\beta(\hat{\mathbf{x}}, t)$ is a function that modulates the potential according to the perturbations. For fixed coordinate $\mathbf{x}_c$ and time t, the Hamiltonian $\hat{H}_c = \hat{H}[\hat{\mathbf{x}}, \hat{\mathbf{k}}, \beta(\mathbf{x}_c, t)]$ is diagonalized by a local basis of Bloch states. We take only the lattice periodic part of the Bloch states, which gives us the local basis $\{|u_n(\mathbf{k}, \mathbf{x}_c, t)\rangle\}_n$, with eigenenergies $E_n(\mathbf{k}, \mathbf{x}_c, t)$. We are interested in some (nearly) degenerate subset of these bands, which is well separated from the other bands, i.e. the local perturbations are small compared to the gap. Since the wave packet is assumed to stay within the nearly degenerate band subset, we may construct it as

$$\langle \mathbf{r}|\Psi\rangle = \int d\mathbf{k}\, a(\mathbf{k}, t) e^{i\mathbf{k}\mathbf{r}} \sum_n \eta_n(k, t)|u_n(\mathbf{k}, \mathbf{x}_c, t)\rangle. \tag{4.16}$$

The function $a(\mathbf{k}, t)$ defines the overall shape of the wave function, is normalized as $\int d\mathbf{k}|a(\mathbf{k}, t)|^2 = 1$, and is localized at $\mathbf{k}_c$, with a spread that is large enough to still allow for localizing the wave packet position at $\mathbf{x}_c$, i.e. the spread in real space is large compared to the underlying lattice constant. The vector components $\eta_n(\mathbf{k}, t)$ parametrize the spinor motion in the nearly degenerate subspace, with normalization condition $\eta^\dagger \eta = 1$. The effective Hamiltonian, linearised at the position $\mathbf{x}_c$ of the wave packet, is

$$\tilde{H}_c = \hat{H}_c + \frac{1}{2}\{\frac{\partial \hat{H}_c}{\partial \mathbf{x}_c}, \hat{\mathbf{x}} - \mathbf{x}_c\}, \tag{4.17}$$

where the second term is the gradient correction, and this Hamiltonian includes time dependency due to the motion of $\mathbf{x}_c(t)$. The center coordinate is given by

$$\mathbf{x}_c = \langle \Psi|\hat{\mathbf{x}}|\Psi\rangle \approx \frac{\partial \Gamma_c}{\partial \mathbf{k}_c} + i\langle \psi|_c \frac{\partial}{\partial \mathbf{k}_c}|\psi\rangle_c \tag{4.18}$$

where the complex phase Γ has been introduced by $a = e^{-i\Gamma}|a|$, and $|\psi\rangle = \sum_n \eta_n |u_n\rangle$, and the subscript c stands for evaluation at center coordinates.

The calculations of Niu and Sundaram [Sun+99] and Culcer [Cul+05] are based on a variational approach to obtain the semiclassical Lagrangian. The exact Schrödinger equation may be

obtained by variation of the Lagrangian $L = \langle\psi|i\partial_t - \hat{H}|\psi\rangle$ after the set of $\psi^*(\mathbf{x})$ for all $\mathbf{x}$, regarding the wave function at different coordinates as independent functions of time. If one adds some energy shift $E(t)$ to the Hamiltonian, there is an ambiguity since the resulting Schrödinger equation will be different depending on whether the normalization condition $\langle\psi(t)|\psi(t)\rangle = 1$ is used before or after the variational calculation. However this changes only the unimportant dynamical phase, and not the Berry phase. The idea is to reduce the set of variation parameters. If one already knows a subspace where a good approximation to the solution is expected, one may choose a smaller set of parameters parametrizing wave functions of this subset, and thus find an effective Lagrangian of these parameters. In the case of the wave packet, these parameters are the center coordinates $(\mathbf{x}_c, \mathbf{k}_c)$, supplemented by the spinor coefficient vectors $\eta, \eta^\dagger$ in the case of (quasi-)degenerate bands. All of these are regarded as independent functions of the time t, only subject to the constraint $\eta^\dagger\eta = 1$ for the spinor. The variational approach does not make use of the adiabatic theorem. We keep it in mind because we will need wave packet expectation values for the gradient correction.

However, we will take a different path since we want to make contact to the adiabatic theorem. The procedure shown here is not as rigorous as the variational approach, but I find it more transparent since it is symmetrical in the space and momentum center coordinates, and since it also mostly avoids lengthy calculations with wave packets expectation values. After all, quadratic in k terms, corresponding to the spread of the wave packet, are neglected in the works by Niu et. al. anyway. Evaluating the expectation value $\langle\Psi|\tilde{H}_c|\Psi\rangle$ and neglecting terms that are due to the spread, as done by Niu et. al, amounts to substituting $\mathbf{k} \to \mathbf{k}_c$ in the basis states and the energy function. We obtain the time-dependent local Hamiltonian $\mathcal{H}$ for the subspace of the η_n, by taking $E_n(\mathbf{k}_c, \mathbf{x}_c, t)$ as diagonal matrix and adding the gradient correction and also a small perturbation that lifts the strict degeneracy (and does not render the equations more complicated),

$$\mathcal{H}^d_{nm} = \langle u_n|_c(\hat{H}_c(\mathbf{k}_c) - E_{nc})|u_m\rangle_c. \tag{4.19}$$

We use the short notations $E_{nc} = E_n(\mathbf{k}_c, \mathbf{x}_c, t)$ and $|u_n\rangle_c = |u_n(\mathbf{k}_c, \mathbf{x}_c, t)\rangle$. To evaluate the gradient correction in (4.17), we neglect the small lifting of the degeneracy by (4.19), considering its influence as higher order perturbation. Further, we assume that approximately, $\hat{H}_c|\Psi\rangle \approx E_{nc}|\Psi\rangle$ and therefore, $\langle\Psi|\frac{\partial\hat{H}_c}{\partial\mathbf{x}_c}|\Psi\rangle \approx \frac{\partial E_{nc}}{\partial\mathbf{x}_c}$. These relations would be exact if $|\Psi\rangle$ had only contributions of eigenstates with exact energy E_{nc}, i.e. we have neglected the spread of the wave packet. The gradient correction is

$$\Delta E^g = \langle\Psi|\frac{1}{2}\left\{\frac{\partial\hat{H}_c}{\partial\mathbf{x}_c}, \hat{\mathbf{x}} - \mathbf{x}_c\right\}|\Psi\rangle \approx \mathrm{Re}\langle\Psi|\hat{\mathbf{x}}\frac{\partial\hat{H}_c}{\partial\mathbf{x}_c}|\Psi\rangle - \mathbf{x}_c\frac{\partial E_{nc}}{\partial\mathbf{x}_c} \tag{4.20}$$

$$= \mathrm{Im}\langle\nabla_{\mathbf{k}}\Psi|\left[\frac{\partial}{\partial\mathbf{x}_c}, \hat{H}_c\right]|\Psi\rangle - \mathbf{x}_c\frac{\partial E_{nc}}{\partial\mathbf{x}_c} \approx \mathrm{Im}\langle\nabla_{\mathbf{k}}\Psi|(E_{nc} - \hat{H}_{nc})|\nabla_{\mathbf{x}_c}\Psi\rangle, \tag{4.21}$$

where we have made use of $\frac{\partial\hat{H}_c}{\partial\mathbf{x}_c} = [\frac{\partial}{\partial\mathbf{x}_c}, \hat{H}_c]$ and $\mathbf{x}_c = \langle\Psi|\hat{\mathbf{x}}|\Psi\rangle = \mathrm{Im}\langle\nabla_{\mathbf{k}}\Psi|\Psi\rangle$ in the last two steps. For further evaluation of (4.21), we note that any term with derivatives $\nabla_{\mathbf{x}_c}$ or $\nabla_{\mathbf{k}}$ acting on the factors $a(\mathbf{k}, t)e^{i\mathbf{kr}}\eta_n(\mathbf{k}, t)$ in $|\Psi\rangle$ cancels (as seen in (4.18), the phase of $a(\mathbf{k}, t)$ also controls the center coordinate $\mathbf{x}_c$). For a superposition of eigenstates of same energy E_{nc}, $|\psi\rangle = \sum_n \eta_n|u_n\rangle_c$, we obtain the gradient correction $\Delta E^g = \sum_{nm}\eta^*_n\mathcal{H}^g_{nm}\eta_m$ with

$$\mathcal{H}^g_{nm} = \mathrm{Im}\langle\nabla_{\mathbf{k}_c}u_n|_c(E_{nc} - \hat{H}_c(\mathbf{k}_c)|\nabla_{\mathbf{x}_c}u_m\rangle_c. \tag{4.22}$$

So the time dependent local Hamiltonian for the quasi-degenerate subspace is

$$\mathcal{H}_{nm} = \delta_{nm} E_{nc} + \mathcal{H}^d_{nm} + \mathcal{H}^g_{nm}. \tag{4.23}$$

The adiabatic Schrödinger equation for the components η_n of the wave function $|u\rangle = \sum_n \eta_n |u_n\rangle_c$, including the above corrections, reads

$$i\frac{d}{dt}\eta_n = \sum_m \left(\mathcal{H}_{nm} - i\langle u_n|_c \frac{d}{dt}|u_m\rangle_c \right) \eta_m. \tag{4.24}$$

Here, $\mathcal{H}_{nm}$ and the basis states $\{|u_n\rangle\}$ may depend on time both explicitly and through the center coordinates.

Our next goal is to derive the equations of motion for the (classical) coordinates[2] $\mathbf{x}_c$ and $\mathbf{k}_c$. In the non-degenerate case and without any Berry phase corrections, the Lagrangian $\mathcal{L}(\mathbf{x}_c, \dot{\mathbf{x}}_c) = \dot{\mathbf{x}}_c \mathbf{k}_c - E(\mathbf{x}_c, \mathbf{k}_c)$ is obtained from the Legendre transform of the classical Hamiltonian, given by the energy $E(\mathbf{x}_c, \mathbf{k}_c)$. Although the Legendre transform no longer depends on $\mathbf{k}_c$, we may regard the same Lagrangian as function also of $\mathbf{k}_c$ and $\dot{\mathbf{k}}_c$, $\mathcal{L}(\mathbf{x}_c, \mathbf{k}_c, \dot{\mathbf{x}}_c, \dot{\mathbf{k}}_c)$, thereby only adding the trivial equation of motion $\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{k}}_c} = 0 = \frac{\partial \mathcal{L}}{\partial \mathbf{k}_c} = 0$. This enables us to add the Berry connection as velocity-dependent correction to the energy, which would not be possible in the Hamiltonian formalism. This procedure can be justified by the fact that $\frac{\partial \mathcal{L}}{\partial \eta^\dagger} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\eta}^\dagger}$ will reproduce the adiabatic equation of motion (4.24). We now expand the total derivative $\frac{d}{dt}|u_m\rangle_c = (\dot{\mathbf{x}}_c \nabla_{\mathbf{x}_c} + \dot{\mathbf{k}}_c \nabla_{\mathbf{k}_c} + \partial_t)|u_m\rangle_c$, so

$$\mathcal{L}(\mathbf{x}_c, \mathbf{k}_c, \dot{\mathbf{x}}_c, \dot{\mathbf{k}}_c, \eta, \eta^\dagger) = \sum_{nm} \eta^\dagger_n \left(i\frac{d}{dt} + \dot{\mathbf{x}}_c \mathbf{k}_c - \mathcal{H}_{nm} + i\langle u_n|_c (\dot{\mathbf{x}}_c \nabla_{\mathbf{x}_c} + \dot{\mathbf{k}}_c \nabla_{\mathbf{k}_c} + \partial_t)|u_m\rangle_c \right) \eta_m. \tag{4.25}$$

The equations of motion for $\mathbf{x}_c, \mathbf{k}_c$ are obtained from this Lagrangian, in the usual way by the Euler-Lagrange equations $\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{k}}_c} = \frac{\partial \mathcal{L}}{\partial \mathbf{k}_c}$ and $\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}_c} = \frac{\partial \mathcal{L}}{\partial \mathbf{x}_c}$. Introducing the 6×6 matrix $J = \begin{pmatrix} & \mathbb{1} \\ -\mathbb{1} & \end{pmatrix}$, and the vector $\mathbf{s} = (\mathbf{x}_c, \mathbf{k}_c)$, we may write them in a compact form,

$$J_{\alpha\beta}\dot{s_\beta} + \eta^\dagger \frac{\partial \mathcal{H}}{\partial s_\alpha}\eta = -\eta^\dagger \left(F^B_{\alpha\beta}\dot{s_\beta} + F^B_{\alpha t} + i[A^B_\alpha, \mathcal{H}] \right) \eta \tag{4.26}$$

with the generalized, matrix-valued Berry connections

$$(A^B_\alpha)_{nm} = -i\langle u_n|_c \frac{\partial}{\partial R_\alpha}|u_m\rangle_c, \quad (A^B_t)_{nm} = -i\langle u_n|_c \frac{\partial}{\partial t}|u_m\rangle_c. \tag{4.27}$$

Here the indices α, β take values of the six coordinates, and summation over β and the vector components of η is implied in (4.26). Note that A^B_t is defined slightly different from $A^B(t)$, the difference being the derivatives $\frac{\partial}{\partial t}$ and $\frac{d}{dt}$. Further, we have used the generalized Berry curvatures, which include non-Abelian (commutator) terms, compare Eq. (4.15),

$$(F^B_{\alpha\beta})_{nm} = \left(\partial_\alpha A^B_\beta - \partial_\beta A^B_\alpha\right) + i[A^B_\alpha, A^B_\beta] \tag{4.28}$$

$$= -i\left(\langle \partial_\alpha u_n|_c \partial_\beta u_m\rangle_c - \langle \partial_\beta u_n|_c \partial_\alpha u_m\rangle_c\right) + i[A^B_\alpha, A^B_\beta], \tag{4.29}$$

$$(F^B_{\alpha t})_{nm} = -i\left(\langle \partial_\alpha u_n|_c \partial_t u_m\rangle_c - \langle \partial_t u_n|_c \partial_\alpha u_m\rangle_c\right) + i[A^B_\alpha, A^B_t] \tag{4.30}$$

[2] Note that different from Culcer and Niu, we treat $\mathbf{x}$ and $\mathbf{k}$ in a symmetric way, by fixing them both to center coordinates $\mathbf{x}_c$ and $\mathbf{k}_c$. In their case, basis states of fixed $\mathbf{x}_c$ are used, but the wave packet - and also the expansion coefficients η_n - still depend on a set of momenta $\mathbf{k}$ centred around $\mathbf{k}_c$.

The left hand side of (4.26) contains the usual Hamiltonian equations of motion, while the right hand side appears due to the adiabatic motion of the spinor and reveals the structure of a non-Abelian gauge theory. The Hamiltonian equations are invariant under the symplectic symmetry of canonical transformations, i.e. coordinate transformations $\mathbf{s} = (\mathbf{x}_c, \mathbf{k}_c) \to \mathbf{S} = (\mathbf{X}_c, \mathbf{K}_c)$ that have an invertible transformation matrix $W_{ij} = \frac{\partial(\mathbf{x}_c, \mathbf{k}_c)_i}{\partial(\mathbf{X}_c, \mathbf{K}_c)_j}$ which obeys the condition

$$W^T J W = J, \tag{4.31}$$

with the symplectic unitary 6×6 matrix J introduced above. The symplectic transformations form a group. Indeed, both sides of Eq. (4.26) are invariant under this symmetry, showing that it gives a Hamiltonian system of equations. To show the invariance, we introduce the quantities

$$\tilde{E} = \eta^\dagger \mathcal{H} \eta, \quad \tilde{F}_{\alpha\beta} = \eta^\dagger F^B_{\alpha\beta} \eta, \quad \tilde{A}_\alpha = \eta^\dagger \left(F^B_{\alpha t} + i[A^B_\alpha, \mathcal{H}] \right) \eta. \tag{4.32}$$

With these, the semiclassical equations (4.26) can be rewritten in the convenient form

$$\left(J + \tilde{F} \right) \dot{\mathbf{s}} = \nabla_{\mathbf{s}} \tilde{E} + \tilde{A}. \tag{4.33}$$

Now the symplectic invariance becomes apparent by the transformation properties $\nabla_{\mathbf{s}} \to \nabla_{\mathbf{S}} = W^T \nabla_{\mathbf{s}}$, $\tilde{A} \to W^T \tilde{A}$, $\dot{\mathbf{s}} \to \dot{\mathbf{S}} = W^{-1} \dot{\mathbf{s}}$ and $\tilde{F} \to W^T \tilde{F} W$. Note that the symmetry does not trivially follow from the fact that the semiclassical equations are obtained from a Lagrangian, since $\mathbf{x}_c$ and $\mathbf{k}_c$ are not conjugate variables but independent variables in $\mathcal{L}(\mathbf{x}_c, \mathbf{k}_c, \dot{\mathbf{x}}_c, \dot{\mathbf{k}}_c, \eta, \eta^\dagger)$.

As an example, let us apply the semiclassical wave packet theory to a non-degenerate Bloch band. As slowly varying perturbations, we add a homogeneous magnetic field by the vector potential $\mathbf{A}(\mathbf{x}_c) = \frac{1}{2} \mathbf{B} \times \mathbf{x}_c$ and a homogeneous electric field $\mathbf{E} = -\nabla \phi(\mathbf{x}_c)$ by the electrostatic potential $-e\phi(\mathbf{x}_c)$. Then, the local Hamiltonian becomes

$$\hat{H}_c = \frac{(\hat{\mathbf{k}} + e\mathbf{A}(\mathbf{x}_c))^2}{2m_0} + V_0(\hat{\mathbf{x}}) - e\phi(\mathbf{x}_c), \tag{4.34}$$

where $V_0(\hat{\mathbf{x}})$ is the lattice-periodic crystal potential. Since $e\mathbf{A}(\mathbf{x}_c)$ is not treated as an operator but as classical shift of the crystal momentum $\mathbf{k}$, the form of the Bloch states remains unmodified by the magnetic field. The Bloch state at the center coordinates is $|u\rangle_c = |u(\mathbf{k}_c + e\mathbf{A}(\mathbf{x}_c))\rangle$. The equations of motions simplify if we introduce the variable $\mathbf{p}_c = \mathbf{k}_c + e\mathbf{A}(\mathbf{x}_c)$, and treat $(\mathbf{x}_c, \mathbf{p}_c)$ as pair of independent variables. Then, the Berry curvatures (4.29) depend on components of $\mathbf{p}_c$, but no longer on $\mathbf{x}_c$. The semiclassical energy $E_c = E(\mathbf{p}_c) - \phi(\mathbf{x}_c) + E_M(\mathbf{p}_c)$ includes a gradient correction $E_M(\mathbf{p}_c)$ corresponding to the orbital magnetization energy of the Bloch state. Since the system is assumed to be non-degenerate, we put $\eta = \eta^\dagger = 1$. The semiclassical equations of motions read [Sun+99]

$$\dot{\mathbf{x}}_c = \frac{\partial E_c}{\partial \mathbf{p}_c} - \dot{\mathbf{p}}_c \times \mathbf{F}^B, \quad \dot{\mathbf{p}}_c = -e\,\mathbf{E} - e\,\dot{\mathbf{x}}_c \times \mathbf{B} \tag{4.35}$$

where the Berry curvature

$$\mathbf{F}^B = \frac{i}{2} \epsilon_{jkl} \mathbf{e}_l \left(\langle \partial_{p_{cj}} u | \partial_{p_{ck}} u \rangle - \langle \partial_{p_{ck}} u | \partial_{p_{cj}} u \rangle \right) \tag{4.36}$$

takes the place of a reciprocal magnetic field. While a usual magnetic field may depend on $\mathbf{r}$, $\mathbf{F}^B$ depends on the kinematic momentum $\mathbf{p}_c$, i.e. on reciprocal space.

If one is interested in an effective quantum mechanical description that also models the wave packet spreading, or quantization of energy levels in a finite system, one must find some re-quantization procedure for the center coordinates in (4.26). But one has to be careful, in this case, since in the presence of the Berry curvatures, ($\mathbf{x}_c$, $\mathbf{k}_c$) can no longer be identified as canonical pair. Sundaram shows when the only perturbation term is the vector potential of a magnetic field, $\mathbf{k}+e\mathbf{A}$ becomes the gauge invariant crystal momentum and $(\mathbf{x}_c, \mathbf{k}_c)$ can be taken as canonical pair, as usual. In the more general case, Sundaram proposes to use a semiclassical rule for quantization of energy levels, meaning that for closed orbits of the semiclassical theory, a constraint in terms of the Berry phase integrated along the path is used, so that bound states corresponding to classically closed orbits, will exist only for discrete energies. Let us preliminarily conclude that the envelope function method is probably more appropriate to find an effective quantum theory.

The search for a re-quantization procedure of the semiclassical equations is further complicated by the fact that an operator $\hat{\mathbf{x}}_c$ that should represent the center of the wave packet, should actually have a spin structure. In our definition (4.18) of $\mathbf{x}_c$ as an expectation value, a possible spin structure is averaged out. By comparison, the Foldy-Wouthysen transformation [Fol+50] gives a quantum mechanical effective low-energy limit which also allows insight into the spin structure. In the next section, we will compare with this approach.

4.2 Wave packets as method for the non-relativistic limit

It is worth while comparing the wave packet approach to the transformation discussed by Foldy in his seminal paper [Fol+50] for the case of the Dirac equation, in order to obtain an effective theory for the non-relativistic limit. Since there is an one-to-one correspondence between the spin-orbit terms of the Dirac equation and a spin-orbit term that we will encounter in an effective quantum well model in Chapter 7, this discussion is relevant to us. The similarity was also recognized in [Bér+06], where a quantum mechanics of non-commuting space operators is discussed and applied to the spin Hall effect.

Let U be the unitary operator that diagonalizes the unperturbed Hamiltonian, i.e. $U = U(\mathbf{k})$ in the context of the wave packet approach or $U = U(\mathbf{p})$ in the context of the Dirac equation. The Dirac equation for a free relativistic spin-$\frac{1}{2}$ particle of mass m_0 reads (we set $\hbar = 1$)

$$i\frac{\partial \psi}{\partial t} = H_D\psi = \left(\beta m_0 c^2 + c\boldsymbol{\alpha}\cdot\mathbf{p}\right)\psi \tag{4.37}$$

with 4×4 matrices $\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}$ and $\beta = \begin{pmatrix} \mathbb{1} & \\ & -\mathbb{1} \end{pmatrix}$. The solutions ψ are 4-component spinors, where in general, at least 3 components are non-zero. Interestingly, the Dirac velocity operator $\frac{\partial H_D}{\partial \mathbf{p}} = c\boldsymbol{\alpha}$ has eigenvalues $\pm c$, each two-fold degenerate. Thus, an instantaneous measurement of the velocity of a free particle will always give the speed of light! If we decompose a state $|\psi\rangle = \sum_{s=\pm} c_s|\alpha_i, s\rangle$ in eigenstates of $c\alpha_i$ belonging to the eigenvalues $\pm c$, the probabilities to obtain the speed $\pm c$ are given by $|c_\pm|^2$. The expectation value is $c(|c_+|^2 - |c_-|^2)$, which in general will be way below the speed of light.

For a useful effective theory, the velocity operator should be a constant of motion for the free particle, and the value should be given by the expectation value $\langle\psi|\frac{\partial H_D}{\partial \mathbf{p}}|\psi\rangle$. The diagonalization of H_D gives $UH_DU^\dagger = \beta E_p$ with $E_p = \sqrt{m_0^2c^4 + p^2c^2}$. So if $|u_\mathbf{p}\rangle$ is an eigenstate of H_D of positive (negative) energy, only the two upper (lower) components of $U|u_\mathbf{p}\rangle$ will be non-zero.

Foldy [Fol+50] finds that the operator

$$\mathbf{X} = U^\dagger \mathbf{x} U = \mathbf{x} + iU^\dagger \frac{\partial U}{\partial \mathbf{p}} = \mathbf{x} + \frac{ic\beta\boldsymbol{\alpha}}{2E_p} - \frac{ic^2\beta(\boldsymbol{\alpha}\cdot\mathbf{p})\mathbf{p} + c^2(\boldsymbol{\sigma}\times\mathbf{p})p}{2E_p(E_p+m_0c^2)p} \tag{4.38}$$

is the correct *mean position* operator if we are looking for a meaningful velocity operator $\dot{\mathbf{X}} = \frac{1}{i}[\mathbf{X}, H_D] = U^\dagger \frac{\partial E_p}{\partial \mathbf{p}} \beta U$ in terms of the canonical position operator $\mathbf{x} = i\nabla_{\mathbf{p}}$. $\dot{\mathbf{X}}$ is diagonal in the same basis as H_D. Although p_i is the generator of translations in x_i and not in X_i, the commutation relations are identical, $[p_i, x_j] = [p_i, X_j] = \frac{1}{i}\delta_{ij}$. We can easily check that the expectation value of $\dot{\mathbf{x}}$ for an eigenstate of positive energy E_p, $\langle u_{\mathbf{p}}|\dot{\mathbf{x}}|u_{\mathbf{p}}\rangle = \frac{\partial E_p}{\partial \mathbf{p}}$, is the same as that of $\dot{\mathbf{X}}$,

$$\langle u_{\mathbf{p}}|\dot{\mathbf{X}}|u_{\mathbf{p}}\rangle = \frac{1}{i}\langle u_{\mathbf{p}}|U^\dagger[\mathbf{x}, \beta E_p]U|u_{\mathbf{p}}\rangle = \frac{\partial E_p}{\partial \mathbf{p}} = \frac{\mathbf{p}c^2}{E_p}. \tag{4.39}$$

For the free particle, instantaneous measurements of $\dot{\mathbf{X}}$ always give the same velocity, $\frac{pc^2}{E_p} \approx \frac{p}{m_0}$ in the non-relativistic limit. Now that we have found that the mean position operator $\mathbf{X}$ is useful for the non-relativistic limit, it is interesting to consider its eigenstates. As shown by Foldy [Fol+50], the eigenstates of $\mathbf{X}$ can be constructed as wave packets consisting of $\mathbf{x}$-eigenstates, and their spread about the center of position is of the order of the Compton wavelength $\lambda_c = \frac{1}{m_0 c}$. The energy sign operator $U^\dagger \beta U$ commutes with $\mathbf{X}$, allowing one to clearly distinguish between wave packets of particles and antiparticles, and they have a non-trivial spin structure. We can identify the particle and antiparticle solutions with energy bands in the language of semiconductors. Since the wave packets consist of only one degenerate band, they are comparable to the semiclassical wave packets of Section 4.1. Switching between the operators $\mathbf{x} \leftrightarrow \mathbf{X}$ amounts to a unitary basis change. While a basis change applied to $\mathbf{x}$ and H_D at the same time has no physical significance, switching between $\mathbf{x}$ and $\mathbf{X}$ while keeping the basis of H_D does have physical significance and allows us to focus on the definite measurement of different physical quantities. $\mathbf{x}$ is the convenient operator when working within real space - e.g. adding up an electrostatic potential term requires a definite space coordinate. In contrast, switching to operators $\mathbf{X}$ and $\dot{\mathbf{X}}$ gives a velocity operator that is a good quantum number for a free particle, while $\mathbf{X}$ no longer corresponds to a definite place.

In order to derive the effective non-relativistic theory including SO terms, which will be the Pauli equation, we follow Foldy [Fol+50] and add up an electrostatic potential term $V(\mathbf{x}) = e\mathbf{E}\cdot\mathbf{x}$. The potential is assumed to be weak compared with the relativistic energy gap $2m_0c^2$, i.e. it should be slowly varying on the length scale given by the wave packet's spread. Here we are only interested in terms of an effective model up to first order in the homogeneous electrical field $\mathbf{E}$.

Let us switch to the basis which diagonalizes the free H_D. The perturbed Dirac Hamiltonian is given by

$$H' = U(H_D + V(\mathbf{x}))U^\dagger = \beta E_{\mathbf{p}} + e\mathbf{E}\cdot U\mathbf{x}U^\dagger \tag{4.40}$$

In this basis, the mean position operator looks simple, $\mathbf{X}' = i\nabla_{\mathbf{p}} = \mathbf{x}$, and $\mathbf{x}' = U\mathbf{x}U^\dagger = \mathbf{X}' + \mathbf{A} + \mathbf{G}$ with a block-diagonal part $\mathbf{A}$ (not to be confused with the vector potential, which we do not include here) and a block-off-diagonal part $\mathbf{G}$ given by

$$\mathbf{A} = -\frac{c^2(\boldsymbol{\sigma}\times\mathbf{p})}{2E_p(E_p+m_0c^2)}, \tag{4.41}$$

$$\mathbf{G} = -\frac{ic\beta\boldsymbol{\alpha}}{2E_p} + \frac{ic^2\beta(\boldsymbol{\alpha}\cdot\mathbf{p})\mathbf{p}}{2E_p(E_p+m_0c^2)p}. \tag{4.42}$$

Further, the projection onto the positive energy band, $P = \sum_{i=1,2} |u_{\mathbf{p},i}\rangle\langle u_{\mathbf{p},i}|$ where the $|u_{\mathbf{p},i}\rangle$ are column vectors of $U^\dagger$, has the simple $\mathbf{p}$-independent representation $P = \mathrm{diag}(1,1,0,0)$ in the primed basis.

The only block-off-diagonal term in H' is $e\mathbf{E}\cdot\mathbf{G}$. In principle, a second unitary basis rotation to $H'' = U_2 H' U_2^\dagger$ can be used to eliminate off-diagonal blocks[3]. The block-diagonalization procedure corresponds to the quasi-degenerate perturbation method explained in Appendix B, which is developed in orders of the inverse mass $1/m_0$. However, from the formulas for perturbation theory (Eq. B.12) it becomes apparent that corrections in the diagonal blocks of H'' due to off-diagonal blocks in H' are of higher order in the field $\mathbf{E}$. Therefore, being interested only in spin-orbit terms linear in $\mathbf{E}$, we may simply ignore off-diagonal blocks, putting them to zero by a projection operation. This observation allows to effectively identify the perturbative block-diagonalization with the operation of projection to the particle band. Being interested in the equations only for the particle band, we may write $PU_2\mathbf{x}'U_2^\dagger P \approx P\mathbf{x}'P$. This is the key point which allows us to make the connection to the theory of wave packets, where the projection to one band is assumed from the beginning. The resulting decoupled positive-energy block gives the Pauli equation

$$E\psi = H_P\psi \equiv \left(E_{\mathbf{p}} + V(\mathbf{x}) - \frac{c^2(\boldsymbol{\sigma}\times\mathbf{p})}{2E_p(E_p+m_0c^2)}\nabla V(\mathbf{x})\right)\psi \tag{4.43}$$

$$\approx \left(m_0c^2 + \frac{p^2}{2m_0} + V(\mathbf{x}) - \frac{e}{4m_0^2c^2}(\boldsymbol{\sigma}\times\mathbf{p})\cdot\mathbf{E}\right)\psi. \tag{4.44}$$

Differently from the Dirac equation, the physical meaning of $\mathbf{x} = i\nabla_{\mathbf{p}}$ here is the mean position. We have expanded up to second order in the small parameter $\frac{\mathbf{p}c}{m_0c^2}$, wherein $\mathbf{p}$ is the derivative operator which also acts on $V(\mathbf{x})$.

In order to compare with the center of coordinate (4.18) in the wave packet approach, we now define a center coordinate operator as the 2×2 matrix

$$\hat{\mathbf{x}}_c = P\mathbf{x}'P = \mathbf{X}' + \mathbf{A}^+, \tag{4.45}$$

where we have introduced the Berry connection $\mathbf{A}^+ = P\mathbf{A}P$ of the particle band,

$$(A_j^+)_{nm} = (A_j)_{nm} = i\langle u_{\mathbf{p},n}|\frac{\partial}{\partial p_j}|u_{\mathbf{p},m}\rangle \quad n,m\in\{1,2\}. \tag{4.46}$$

Due to the Berry connection that appears as correction, $\hat{\mathbf{x}}_c$ acquires a spin structure, and it is non-commutative, i.e. $[\hat{x}_{ci},\hat{x}_{cj}] \neq 0$ for $i\neq j$. The non-commutativity appears only because of the projection after the unitary basis transformation U. A quantum mechanics with such non-commuting space operators has been used in the context of the spin Hall effect [Bér+06], and also for wave packets applied to the anomalous Hall effect [Sin08]. $\hat{\mathbf{x}}_c$ should not be seen as a physical quantity, since it is not gauge invariant due to the Berry connection term, but rather as a tool for calculations that should be consistent up to first order in $\mathbf{E}$ [Sin08].

Now we are able to calculate the anomalous velocity, which is a spin-dependent (i.e. 2×2 - matrix valued) correction to the ordinary velocity, by using the Pauli Hamiltonian H_P as defined in (4.43),

$$v_j^{an} = \frac{1}{i}[\hat{x}_{c,j}, H_P] - \frac{\partial E_p}{\partial p_j} = \sum_l \left(\partial_{p_j}A_l^+ - \partial_{p_l}A_j^+ + \frac{1}{i}[A_j^+, A_l^+]\right)eE_l = \sum_j F^{jl}eE_l \tag{4.47}$$

[3] Note that Foldy discusses the unitary transformation a bit differently, combining both U and U_2 into one block-diagonalization procedure, which however involves more steps.

and we have employed the generalized Berry curvature, which is defined on the space of the positive energy band,

$$F^{jl} = \partial_{p_j} A_l^+ - \partial_{p_l} A_j^+ + \frac{1}{i}[A_j^+, A_l^+]. \tag{4.48}$$

It is defined just as (4.15) of the wave packet approach, with the Berry connection $\mathbf{A}^+$ given by (4.46), and it includes a non-Abelian (commutator) term because the particle band is degenerate.

To obtain an explicit result showing that the anomalous velocity is indeed spin dependent and transverse with respect to the accelerating field, we use (4.41) and we expand up to second order in the inverse energy gap. The parts $\partial_{p_j} A_l^+ eE_l$ and $-\partial_{p_l} A_j^+ eE_l$ in (4.47) give identical contributions, while the commutator part is of higher order in $1/m_0$. We obtain

$$v_j^{an} \approx \frac{e}{2m_0^2 c^2}(\boldsymbol{\sigma} \times \mathbf{E})_j. \tag{4.49}$$

If we leave out the non-commuting correction in $\hat{x}_{c,j}$, and calculate an anomalous velocity from the expression $v_j'^{an} = \frac{1}{i}[x_j, H_P]$, we obtain only half the value of v_j^{an}. We should keep in mind that x_j in the Pauli equation physically corresponds to the mean coordinate operator, while $\hat{x}_{cj}$ is just a projected version of the position operator of the Dirac equation.[4] It is not surprising that the resulting spin-dependent anomalous velocity is different, if evaluated for an effective operator that includes a momentum-dependent rotation in spin space. In Section 9.4.1 we use the anomalous velocity $v_j'^{an}$ for a qualitative, but not for a quantitative explanation of spin physics. Note also that the Pauli equation and $\hat{\mathbf{x}}_c$ are quantum mechanical objects, so there is no need to restrict ourselves to a regime where quantum mechanical waves behave like particles, in order to derive $\mathbf{v}^{an}$. In particular, the restriction $\lambda_F < \lambda_c$ (Fermi wavelength small compared with Compton wavelength) which is needed to construct wave packets, is not essential to obtain $\mathbf{v}^{an}$. The wave packet formalism is here used for interpretation, but not for derivation.

It is interesting to compare with the velocity, calculated within the full Dirac equation. Consider a solution $|u_{\mathbf{p}}\rangle$ of the free Dirac Hamiltonian H_D. While the spin vector $\boldsymbol{\sigma}$ is not conserved, its projection, given by the helicity operator $\boldsymbol{\sigma} \cdot \mathbf{p}$, still commutes with H_D. We find the velocity $\mathbf{v} = \frac{\mathbf{p}c^2}{E_p}$ and $\dot{\mathbf{p}} = -\nabla V$, independent of spin or helicity, although the latter is a good quantum number of the free particle. This seems to be in contradiction to the spin-dependent transverse anomalous velocity that we have obtained from the Pauli equation. We can understand the difference by recognizing that the result $\mathbf{v} = \frac{\mathbf{p}c^2}{E_p}$ assumes a state of positive energy. However, the potential $V(\mathbf{x})$ induces particle-antiparticle oscillations, since $[U^\dagger \beta U, V(\mathbf{x})] \neq 0$. These oscillations cancel the anomalous (spin-dependent) velocity, on average. If the potential is weak and thereby induces only slow particle-antiparticle oscillations, it will be reasonable to neglect the oscillations on the time scale of interest - this corresponds to the projection to the positive-energy band, in the wave packet approach, and this also allows to observe the spin-dependent anomalous velocity. We conclude that the anomalous velocity can be expressed by the difference of commutators with and without intermediate projection,

$$v_i^{an} = \frac{1}{i}[Px_i'P, PH'P] - \frac{1}{i}P[x_i', H']P = \frac{1}{i}\left(PH'\bar{P}x_i'P - Px_i'\bar{P}H'P\right) \tag{4.50}$$

$$= \frac{e}{i}\sum_j E_j \left(Px_j'\bar{P}x_i'P - Px_i'\bar{P}x_j'P\right) \tag{4.51}$$

[4] The nomenclature chosen here is not intuitive because the words "center" and "mean" have similar meaning, which could lead to incorrect conclusions.

with $\bar{P} = 1 - P$, the projection to the negative energy band. The previous expression $v_i^{an} = \sum_j F^{ij} e E_j$ can be shown by using $U(\partial_{p_i} U^\dagger) = -(\partial_{p_i} U) U^\dagger$.

We may interpret the factor two by which v_j^{an} is greater than $v_j'^{an}$, as fingerprint of the relativistic Thomas factor. The latter is known as the factor by which the SO term in the Pauli equation is smaller than the term obtained by a classical argument, that relates the SO coupling to the effective local magnetic field in the frame of a particle moving through an electric field. To lowest order, a particle of velocity $\mathbf{v}$ moving in a constant field $\mathbf{E}$ sees a magnetic field $-\frac{1}{c^2}\mathbf{v} \times \mathbf{E}$ in its rest frame. If one combines this with the Zeeman interaction $\frac{e}{2m_0}\boldsymbol{\sigma} \cdot \mathbf{B}$ and approximates $\mathbf{v} \approx \frac{\mathbf{p}}{m_0}$, one obtains a SO interaction term $-\frac{e}{2m_0^2 c^2}\boldsymbol{\sigma}(\mathbf{E} \times \mathbf{p})$, which is twice the value of the correct SO term in the Pauli equation (4.43). The reason for this mismatch is that the new reference system, e.g. of an electron circulating a nucleus, is not an inertial system since it is rotating and thus accelerated. The transverse acceleration can be seen as series of infinitesimally small non-collinear boosts. However non-collinear boosts do not commute. If the order of applying the boosts is changed, the resulting velocity differs by an angle, but not in magnitude. It is possible to construct a space of rapidity $\eta_i = \tanh^{-1} \frac{v_i}{c}$, which is conformal, meaning although it is a mathematical space with a non-trivial metric, the transformation to rapidity still conserves angles. In two dimensions, rapidity space consists of a unit circle - see [Rho+04] for a nice introduction. The boosts correspond to trajectories in rapidity space lying on the geodesics (these are the "most straight" lines in a curved manifold, and may be obtained by parallel transport). The geodesics are either straight lines intersecting the origin, or circles intersecting the circle with unit radius at perpendicular angles, with their origin lying outside the circle of the rapidity space. In a situation where a particle of velocity $\mathbf{v}$ is subject to an acceleration $\mathbf{a}$, this additional rotation angle of the local reference frame is called Thomas-Wigner precession angle χ, and can be found from

$$\frac{d\chi}{dt} = \frac{\gamma - 1}{v^2}(\mathbf{a} \times \mathbf{v})_z, \tag{4.52}$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$. The direction of rotation is opposite to the orbital rotation (rapidity space has a hyperbolic geometry). That is, a clockwise rotation of a particle about a center of attraction will result in the reference frame of the particle being rotated anti-clockwise due to parallel transport in rapidity space. We may approximate $\mathbf{a} = -\frac{1}{m_0}\nabla V$, $\gamma \approx 1 + \frac{v^2}{2c^2}$ and $\mathbf{v} \approx \frac{\mathbf{p}}{m_0}$, in order to construct a perturbing Hamiltonian H_{TW} that generates the Thomas-Wigner precession in the spin space, by demanding $e^{-iH_{TW}t} = e^{i\frac{1}{2}\sigma_z \frac{d\chi}{dt}}$. The result is $-\frac{1}{2}$ the interaction of the classical argument,

$$H_{TW} = \frac{e}{4m_0^2 c^2}(\mathbf{E} \times \mathbf{p}) \cdot \boldsymbol{\sigma} \tag{4.53}$$

where we have exploited the fact that the plane of ∇V and $\mathbf{p}$ is normal to the z direction.

Let us make some connection to the Berry connection that can appear as correction in the position operator, depending on the basis. The Pauli equation written in the mean position operator $\mathbf{X}' = \mathbf{x}$ includes energy corrections due to the spin structure of a wave packet in the vicinity of the center of position, so the Thomas-Wigner rotation of the reference frame should enter. On the other hand, since the operator $\hat{\mathbf{x}}_c$ is just a projected version of the position operator in the Dirac equation, it should not capture the spin structure of the wave packet, resulting in a spin-dependent velocity corresponding to the classical argument without the Thomas-Wigner precession, which is larger.

The difficulty of this discussion is, that all contributions to the SO coupling are of relativistic origin, in the end - the Zeeman coupling used for the classical argument is also of relativistic origin. Therefore, just looking at orders in the relativistic gap $2m_0c^2$, it is impossible to regroup terms into purely kinematic relativistic terms (i.e. the Thomas precession) and other relativistic effects.

Chapter 5

Theory of electronic transport

5.1 The Landauer-Büttiker and non-equilibrium Green's function formalisms

In this section, we want to give a discussion of the theory of electronic transport, as far as it is needed for further calculations. Rather than to include all details of the general derivation of the formalism, our aim is here to discuss the physics behind the needed equations. There is already excellent literature available about the Landauer-Büttiker (LB) and Keldysh [Kel65] formalisms both on introductory [Dat07; Dat05] and advanced [Ram07] levels.

In the real (non-idealized) world, transport phenomena are always linked to dissipation. However, the basic equations describing the dynamics of an isolated physical system - be it the Newton equation for a semiclassical description or the Schrödinger equation for the quantum dynamics - are time reversal invariant and therefore, do not allow for dissipation. Thus, while they are suitable to describe the dynamical behaviour of a system, they cannot describe the relaxation to the thermodynamic ground state. Dissipation appears only when the system is coupled to an environment. On a microscopic level, dissipation can be viewed as the loss of information (in terms of knowledge of the system's particle trajectories) that appears due to the coupling with states of the environment, the latter not being under full control of the observer. The simplest and commonly used approach is to assume that the environment has sufficiently good communication of states so that it can be modelled as a heat bath, which stays in thermodynamic equilibrium at all times.

We divide the open system into a finite mesoscopic part (called sample) attached to semi-infinite leads. The leads function as electrical contacts and as thermodynamic reservoirs at the same time. In the case of the LB formalism, dissipative interactions are not present in the sample, and it is this separation into a part purely described by the dynamical equation and a part where only dissipation takes place, which makes the formalism so simple. On the other hand, if interactions are present in the sample (beyond elastic e-e scattering, e.g. phonons), dissipation and dynamics are intertwined, and the Keldysh formalism provides a way to treat such problems.

In this thesis, results are entirely obtained within the regime of linear response, and for single-particle models, where the Landauer-Büttiker formalism applies. In the case of a sample without interactions, the LB formalism can be shown to be equivalent to the non-equilibrium Green's function (NEGF) or Keldysh formalism.

As first recognized by Landauer, the conductance and transmission properties of scattering

processes of a two-terminal device are equivalent, as long as only elastic scattering takes place. By phenomenological and symmetry arguments, the scattering description was later extended by Büttiker [Büt+85] to the treatment of devices with multiple leads.

The very basic setup to be considered is as follows. There is a central scattering region (sample), with some attached wires (leads) which extend to infinity. We divide the Hamiltonian into a sample part, and parts for the leads, here exemplified by two leads p and q, but the generalization to more leads is straightforward. The Hamiltonian of the isolated parts is $\hat{H}_S$ for the sample and $\hat{H}_p$, $\hat{H}_q$ for the leads. We define $\hat{H}_0 = \hat{H}_S + \hat{H}_p + \hat{H}_q$. The coupling terms between sample and lead p and q are given by $\hat{\tau}_p$ and $\hat{\tau}_q$, respectively. We assume that there is no direct coupling between leads, i.e. $\hat{\tau}_p \hat{\tau}_q^\dagger = 0$. The sum of all coupling terms is treated as perturbation in the Keldysh formalism and is given by $\hat{H}' = \hat{\tau}_p + \hat{\tau}_p^\dagger + \hat{\tau}_q + \hat{\tau}_q^\dagger$.

To develop the formalism, it is assumed that up to some time t_0 in the far past, the system is decoupled and in thermodynamic equilibrium. The coupling term is made time dependent by substituting $\hat{H}' \to \theta(t-t_0)\hat{H}'$. However, in the resulting equations, this time dependency will not be of significance, since we are only interested in the steady state regime. Then, initial oscillations of the currents will have decayed and they will be constant. The individual terms may be written in a basis of lattice sites, of a lattice extending over all parts, by $\hat{H}_S = \sum_{i\in S, j\in S} c_i^\dagger c_j H_{S,ij}$, $\hat{H}_p = \sum_{i\in p, j\in p} c_i^\dagger c_j H_{p,ij}$, and $\hat{\tau}_p = \sum_{i\in S, j\in p} c_i^\dagger c_j \tau_{p,ij}$, and analogue definitions for lead q. Here, c_j and $c_j^\dagger$ are the fermionic annihilation and creation operators of electrons on lattice sites j. The sums over $j \in S$, $j \in p$ refer to the subset of lattice sites belonging to the sample or lead-p region. In matrix representation, we may write

$$\hat{H} = \hat{H}_0 + \hat{H}' = \begin{pmatrix} H_S & \tau_p & \tau_q \\ \tau_p^\dagger & H_p & \\ \tau_q^\dagger & & H_q \end{pmatrix}. \tag{5.1}$$

In order to avoid clobbering of the notation, we will leave away the hat over the Hamiltonian parts in the following, thus referring to a submatrix and a full matrix with only one non-zero submatrix by the same symbol, as long at the meaning of the symbol is clear from the context.

The leads are treated as thermodynamic reservoirs with (in general differing) chemical potentials μ_l and temperatures T_l, where the index l denotes the lead. Note that regarding the leads as thermodynamic reservoirs is a simplified view. Since the leads are assumed to be very narrow (typically, their width is tens till hundreds of nm), they also have a low density of states and thus, removing charge carriers from them as in the case of a current, must disturb their equilibrium significantly. Therefore, Datta [Dat07] introduces the concept of electrical contacts that are in turn attached to the leads. The contacts are wide and have such a large density of states that even when current is flowing, they may be assumed to be in thermodynamic equilibrium, with potentials μ_l and inverse temperatures $\beta_l = \frac{1}{k_B T_l}$. The trick is that the interface between contacts and leads is reflectionless in the sense that waves leaving the lead and entering the contact will not be reflected (while this is not the case for the other direction). A nice analysis of this situation, i.e. a wire having an abrupt change of its width, is given in [Sza+89], and this analysis indeed confirms that one may assume contacts to be reflectionless even when the width changes abruptly (if the width changes slowly, the situation is even better).

Therefore, in the following we can make the assumption that ingoing states of a lead, i.e. states propagating from contact to sample, are populated according to the Fermi distribution determined by the potential μ_l, while the outgoing states, i.e. states coming from the sample,

are not necessarily populated according to a simple distribution function, but enter the contact without reflection.

In real samples, most of the time it is not necessary to distinguish between leads and contacts. There will be electrical contacts which are long compared to the dimension of the sample, and they widen slowly up to a macroscopic scale.

5.2 Green's functions

The main idea of the NEGF/Keldysh formalism is to calculate non-equilibrium properties, like e.g. currents, by a time dependent perturbation theory that starts from some initial equilibrium distribution. To be specific, in our setup we imagine that at some time t_0 in the past, the sample is decoupled from the leads (reservoirs). The coupling H' is turned on at t_0, and in our problem, this will be the only time dependency of the full Hamiltonian. After the coupling is turned on, a current begins to flow from the higher voltage reservoir to the lower voltage reservoir. If the reservoirs are infinitely large, a steady state will be reached after some time. Otherwise, the system will have a return time, so a time scale with properties similar to the steady state will be difficult to define. Therefore, if the calculus requires a finite system, a limit of infinite system size is usually carried out, and this step must be performed before taking the limit of late observation times. We will only be interested in the steady state property, since the time dependency by turning on H' is artificial. If the reservoirs are given by reflectionless contacts as in the LB formalism, the problem of time dependent turning on of H' is avoided completely, since the distribution function of ingoing states (i.e. states propagating from lead to sample) is not affected by the states entering the leads and thus always stays in equilibrium.

In principle, the correlation functions that are used in the NEGF formalism, depend on two times t and t'. However, if the Hamiltonian is time independent, they only depend on the difference of the two times. Further, even if the Hamiltonian is time dependent because of the turning on of some interaction or coupling at t_0, if only times are considered where the steady state has been reached, the correlation functions also depend only on the time difference. In this case, it is convenient to use the energy representation obtained by Fourier transformation with respect to $t - t'$, since integrals over times then become ordinary products.

The term *Green's function* originates from the theory of linear differential equations, where a Green's function is defined as solution of the differential equation for a delta-function inhomogeneity. Here and in the literature about the NEGF formalism [Ram07; Eco06], the term Green's function is used in a wider sense, for all kinds of correlation functions. Only the retarded and advanced Green's functions in the time domain really satisfy the mentioned definition, the other correlation functions are called Green's functions by analogy. In the time domain, the most important Green's functions that we need are

$$G^R_{ij}(t,t') = -\frac{i}{\hbar}\theta(t-t')\langle\{c_i(t), c_j^\dagger(t')\}\rangle \tag{5.2}$$

$$G^A_{ij}(t,t') = \frac{i}{\hbar}\theta(t'-t)\langle\{c_i(t), c_j^\dagger(t')\}\rangle \tag{5.3}$$

$$G^<_{ij}(t,t') = \frac{i}{\hbar}\langle c_j^\dagger(t')c_i(t)\rangle \tag{5.4}$$

$$G^>_{ij}(t,t') = -\frac{i}{\hbar}\langle c_i(t)c_j^\dagger(t')\rangle. \tag{5.5}$$

Here, $c_i(t)$ and $c_j^\dagger(t)$ are fermionic annihilation and creation operators of some single-particle basis states, and $\{,\}$ stands for the anticommutator. The usual fermionic anticommutation relations are assumed,

$$\{c_i, c_j\} = 0, \quad \{c_i^\dagger, c_j^\dagger\} = 0, \quad \{c_i, c_j^\dagger\} = \delta_{ij}. \tag{5.6}$$

In our case, it is particularly useful to employ basis functions localized on single sites of a lattice, since for numerical purposes the Hamiltonian of the setup will be formulated on a lattice. But the equations can be equally well written in the field operators $\psi(\mathbf{x})$, $\psi^\dagger(\mathbf{x})$ or in annihilation/creation operators of momentum eigenstates, $c_\mathbf{k}$, $c_\mathbf{k}^\dagger$. If a basis including a spin or band index is needed, the annihilation/creation operators should be $c_{i,\sigma}$, $c_{i,\sigma}^\dagger$ with some extra band index σ. For better readability we will absorb these degrees of freedom in the index i.

The brackets $\langle ... \rangle$ in (5.2) denote the thermodynamic expectation value, for a (equilibrium) density matrix at time t_0, $\rho(t_0) = \frac{1}{Z} e^{-\sum_l \beta_l (H(t_0) - \mu_l)}$ with $Z = \mathrm{Tr}\rho(t_0)$. At time t_0, Heisenberg and Schrödinger pictures coincide and each decoupled part l of the system is in equilibrium. The time dependency of annihilation and creation operators is defined in the Heisenberg picture,

$$c_j^{(\dagger)}(t) = U(t, t_0) c_j^{(\dagger)} U^\dagger(t, t_0) \tag{5.7}$$

$$U(t, t_0) = \hat{T} e^{-\frac{i}{\hbar}\int_{t_0}^{t} dt' H(t')} \tag{5.8}$$

where $\hat{T}$ is the time ordering operator, which shifts the latest times to the left.

G^R and G^A are called retarded and advanced Green's functions, since they can be interpreted as the impulse response to a delta-like perturbation of the Schrödinger equation. The advanced Green's function describes the (unphysical) solution, where the wave moves toward the point of perturbation, i.e. the response is in advance of the perturbation. In the non-interacting case, i.e. when then Hamiltonian is quadratic in the field operators, both $G^{R/A}$ satisfy the differential equation

$$(i\hbar\partial_t - H)\, G^{R/A}(t, t') = \delta(t - t')\mathbb{1}. \tag{5.9}$$

Then, knowing the time evolution $U(t, t')$ is equivalent to knowing G^R and G^A, and we may write $G^R(t, t') = \frac{1}{i\hbar}\theta(t - t')U(t, t')$ and $G^A(t, t') = \frac{i}{\hbar}\theta(t' - t)U(t, t')$. Here and in the following, without indices i, j of $G^R_{ij}(t, t')$, an equation is to be understood as operator or matrix equation. In case interactions are present, like e.g. a Coulomb term $\int\int d\mathbf{r}\, d\mathbf{r}'\, \psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}')V(\mathbf{r}-\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r})$ involving four field operators, the Green's functions obey a hierarchy of differential equations, where single-particle Green's functions like the ones defined above are linked to two-particle Green's functions involving two pairs of creation and annihilation operators, and n-particle Green's functions are linked to $(n+1)$-particle Green's functions [Eco06].

$G^<(t, t')$ and $G^>(t, t')$ are called lesser and greater Green's functions, where the nomenclature originates from the Keldysh formalism. In the Keldysh formalism, a closed contour reaching first from $-\infty$ to ∞ and then back from ∞ to $-\infty$ is introduced for the times t, t'. A generalized time-ordered Green's function depending on two times on this contour is introduced, which unifies all needed Green's functions in one object. Specializing to the case where t lies on the forward axis and t' on the backward axis, one obtains $G^<(t, t')$. On the Keldysh contour, it is then understood that $t \overset{c}{<} t'$ independent of the values $t, t' \in \mathbb{R}$, thus the name "lesser". On the other hand, if t' lies on the forward axis and t on the backward axis, time ordering of the

fermionic field operators on the Keldysh contour gives a minus sign, with explains the definition (5.5) of $G^{>}(t,t')$. In the non-interacting case, $G^{<}$ and $G^{>}$ obey the differential equation

$$(i\hbar\partial_t - H)\, G^{</>}(t,t') = 0. \tag{5.10}$$

Physically, $-i\hbar G^{<}_{ij}(t,t')$ can be understood as the statistical expectation value of the probability amplitude, for finding an electron in state i at time t and in state j at time t'. The same interpretation, but for holes instead of electrons, applies for $i\hbar G^{>}_{ij}(t,t')$. In this chapter, by holes we mean missing electrons and do not refer to the type of the energy band. $G^{<}$ and $G^{>}$ can be understood as generalizations of a density matrix for electrons/holes. For equal times $t = t'$, $-i\hbar G^{<}(t,t)$ is a density matrix, and the diagonal entries $-i\hbar G^{<}_{jj}(t,t)$ simply give the electron density in state $c_i^\dagger|0\rangle$, $|0\rangle$ being the vacuum (or empty band) state. Noting that $G^R(t,t') = \theta(t-t')(G^{>}(t,t') - G^{<}(t,t'))$, we again recognize the interpretation of G^R given earlier. If $t > t'$, the probability amplitude for the propagation of an excitation, regardless of particle or hole, from site j at time t' to site i at time t is given by $G^R_{ij}(t,t')$.

We will use $G^{<}$ as a basis to express all possible (single-particle) observables that we need to calculate. On the other hand, G^R will be needed for the Fisher-Lee relation which relates conductance to the propagation of waves in the sample.

Since we are interested in the steady-state regime, where all the correlation functions only depend on the time difference $t - t'$, is is useful to work with energy representations of Green's functions like

$$G^R(E) = \frac{1}{\hbar}\int_{-\infty}^{\infty} dt\, e^{iE(t-t')/\hbar} G^R(t-t'). \tag{5.11}$$

Then, integrals over intermediate times become simple products. In the following, we will avoid overloaded notation by not always putting the argument (E). However, if we refer to the time domain, we will always put time arguments.

The spectral density is defined as

$$A(E) = iG^{>}(E) - iG^{<}(E) = iG^R(E) - iG^A(E). \tag{5.12}$$

It describes the available states of the system, independent of their occupation, since it is the sum of occupied and unoccupied states. In the case of no interactions, $G^{R/A}(E)$ and thus also $A(E)$ depend on the Hamiltonian, but not on the occupation of the states. This can be seen in the spectral representation

$$G^R(E) = \lim_{\eta\to 0^+} (E + i\eta - H)^{-1}, \quad G^A(E) = \lim_{\eta\to 0^+} (E - i\eta - H)^{-1} = (G^R)^\dagger. \tag{5.13}$$

These relations can be proven with the representation of the Heaviside function

$$\theta(t) = \frac{-1}{2\pi i}\lim_{\eta\to 0^+}\int dE\, e^{-iEt/\hbar}\frac{1}{E+i\eta}. \tag{5.14}$$

In equilibrium (assuming a single chemical potential and temperature, for simplicity), the electron and hole occupation is given by the Fermi function $f_0(E) = (1 + e^{\beta(E-\mu)})^{-1}$,

$$G^{<}(E) = if_0(E)A(E) \tag{5.15}$$

$$G^{>}(E) = -i(1 - f_0(E))A(E). \tag{5.16}$$

This intuitive result even holds when interactions are present [Wim08; Dat07].

5.3 Interpretation of $G^<$ and the Fermi sea

In a single particle picture, the correspondence $\psi_i(t) \to c_i(t)$, $\psi_j^*(t') \to c_j^\dagger(t')$ between single particle wave amplitudes at lattice sites i and annihilation/creation operators at these sites, can be used to find the statistical expectation value of correlation functions. In the notation of second quantization,

$$\psi_i(t)\psi_j^*(t') \to -i\hbar G_{ij}^<(t,t') = \langle c_j^\dagger(t')c_i(t)\rangle. \tag{5.17}$$

For equal times $t = t'$, this rule means that we are replacing the pure state density matrix $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$ by a mixed state density matrix which describes the statistical ensemble. If the statistical ensemble consists of a pure state with a single electron, $|\psi(t)\rangle = \sum_i \psi_i(t)c_i^\dagger|0\rangle$, we use $\rho(t_0)$ to evaluate the right side of (5.17) as

$$\langle c_j^\dagger(t')c_i(t)\rangle = \mathrm{Tr}[\rho(t_0)c_j^\dagger(t')c_i(t)] = \langle\psi(t)|c_j^\dagger U(t',t)c_i|\psi(t)\rangle = \psi_i(t)\psi_j^*(t') \tag{5.18}$$

showing that we can put an equal sign instead of $\to$ in (5.17). Here we assumed that at time t_0, the Heisenberg and Schrödinger pictures coincide, and used $\langle 0|U(t',t)|0\rangle = 1$. $|0\rangle$ denotes the vacuum or empty band state. Further, if $\rho(t_0)$ is a mixed single-particle state, we can replace the left side of (5.17) by a statistical average over pure states, corresponding to a decomposition $\rho(t_0) = \sum_n p_n\rho_n(t_0)$ in terms of pure state density matrices $\rho_n(t_0)$ and probabilities p_n. In this sense, the substitution rule (5.17) is still exact.

However, in this section we want to focus on the many-particle aspect of $G^<$ and not on the statistical average - the latter being easy to understand. In a realistic semiconductor or metal system, all states up to the Fermi energy E_F are already occupied, resulting in the Fermi sea $|FS\rangle$ as ground state, and the interesting physics will consist of excitations to this state rather than to the vacuum or empty band state. Therefore, in the following we want to analyse the propagation of a general single particle excitation in the Fermi sea at time t_0,

$$|\psi\rangle = |\psi(t_0)\rangle = \frac{1}{\sqrt{N_\psi}}\sum_l \psi_l c_l^\dagger|FS\rangle \tag{5.19}$$

by analysing expectation values of $c_j^\dagger(t')c_i(t)$. Since the lattice sites are already partly occupied by the Fermi sea and the Pauli principle must hold, the operator $c_l^\dagger$ will not always create an electron on site l. Instead, sometimes it will destroy the state, and we expect to see some oscillations of the density in the vicinity of site l. The oscillations will strongly depend on E_F. We use the normalization convention $\sum_l |\psi_l|^2 = 1$. Due to the Pauli principle, we also need the extra normalization constant N_ψ (as opposed to the vacuum case). It will depend both on the coefficients ψ_l and the Fermi sea occupation. Note that compared to the single particle problem, the anticommutation relations (5.6) add new physical meaning to the lattice. In the many-body problem, each site can be occupied by exactly one electron (if there is no spin). The maximal possible electron density depends on the number of lattice sites per area. The lattice that we employ for discretization of a one particle Hamiltonian should not be interpreted in this way - it is just a lattice for one electron, and is used to track the electron position, but not the number of electrons.

It is convenient to work with the annihilation and creation operators of the orthonormal set of modes $\{\chi^{(k)}\}_k$ in which the (non-interacting) Hamiltonian is diagonal,

$$\hat{H} = \sum_k \epsilon_k a_k^\dagger a_k \quad \text{with} \quad a_k = \sum_i \chi_i^{(k)} c_i, \quad a_k^\dagger = \sum_i \chi_i^{(k)*} c_i^\dagger. \tag{5.20}$$

The completeness relation $\sum_k \chi_i^{(k)*}\chi_j^{(k)} = \delta_{ij}$ can be used to find the representation $c_i = \sum_k \chi_i^{(k)*} a_k$. The new operators also obey the fermionic anticommutation relations

$$\{a_k, a_{k'}\} = 0, \quad \{a_k^\dagger, a_{k'}^\dagger\} = 0, \quad \{a_k, a_{k'}^\dagger\} = \delta_{kk'}, \tag{5.21}$$

and their time dependency is given by $a_k(t) = e^{-i\epsilon_k t/\hbar} a_k$ and $a_k^\dagger(t) = e^{i\epsilon_k t/\hbar} a_k^\dagger$.

We will evaluate the expectation value $\langle\psi|a_k^\dagger(t)a_{k'}(t')|\psi\rangle$ within the representation in modes, $|\psi\rangle = \frac{1}{\sqrt{N_\psi}}\sum_k \psi_k a_k^\dagger|FS\rangle$, corresponding to (5.19). We use the indices i,j,l,l' to denote lattice sites and k, k',q,q' to denote eigenmodes. The calculation is best done within the occupation number representation, where $n_k \in \{0,1\}$ denotes the occupation of a state k in $|FS\rangle$. We obtain

$$\begin{aligned}\langle FS|a_q a_k^\dagger a_{k'} a_{q'}^\dagger|FS\rangle &= \delta_{kk'}\delta_{qq'}(1-n_q) - \langle FS|a_q a_{k'} a_k^\dagger a_{q'}^\dagger|FS\rangle \\ &= \delta_{qk}\delta_{q'k'}(1-n_q)(1-n_{q'}) + \delta_{qq'}\delta_{kk'}(1-n_q)n_k,\end{aligned} \tag{5.22}$$

and for the normalization constant,

$$N_\psi = \sum_q (1-n_q)|\psi_q|^2, \tag{5.23}$$

so N_ψ is a number in the interval $[0,1]$ which gives the fraction of the excitation above the Fermi sea. The statistical ensemble being given by the pure state $|\psi\rangle$, we find

$$-i\hbar G^<_{k'k}(t',t) = \langle\psi|a_k^\dagger(t)a_{k'}(t')|\psi\rangle = \frac{1}{N_\psi}\psi_k^*(t)\psi_{k'}(t')(1-n_k)(1-n_{k'}) + \delta_{kk'}n_k e^{i\epsilon_k(t-t')/\hbar}, \tag{5.24}$$

where we have defined $\psi_k(t) = e^{-i\epsilon_k t/\hbar}\psi_k$. In the first summand of (5.24), the factors $(1-n_k)(1-n_{k'})$ select only those modes of the excitation, which correspond to energies above the Fermi sea. This gives the quickly oscillating response part, which behaves similar to the vacuum case (5.18). The second summand, $\delta_{kk'}n_k e^{i\epsilon_k(t-t')/\hbar}$, is exactly the correlation function $-i\hbar G^<_{k'k}(t',t)$ of the unperturbed Fermi sea. This part gives the slowly oscillating part of the response, corresponding to the low energy modes, which propagate inside the Fermi sea.

We could now go back to the lattice site representation by $\psi_k = \sum_i \psi_i \chi_i^{(k)}$, and enter the lattice representations (5.20) of a_k and $a_k^\dagger$. But it is more instructive to consider a special case. In order to get some insight into the behaviour of the electron density after applying the creation operator at some site i at time $t_0 = 0$, we look at the resulting real space density matrix for a state $|\psi\rangle = \frac{1}{\sqrt{N_\psi}}c_i^\dagger|FS\rangle$, by specializing to $\psi_k = \chi_i^{(k)}$. This situation could be experimentally relevant if an electron is injected from a scanning tunnelling microscope (STM) tip into a lattice site (or atomic orbital). The injection into a Fermi sea of $N_{occ} = \sum_q n_q$ electrons should result in a state with $N_{occ}+1$ electrons, even if the extra electron cannot be found at site i. The real space representation of Eq. (5.24) becomes

$$-i\hbar G^<_{lj}(t,t) = \frac{1}{N_\psi}\sum_{kk'}\chi_j^{(k)}\chi_l^{(k')*}(1-n_k)(1-n_{k'})\chi_i^{(k)*}\chi_i^{(k')}e^{it(\epsilon_k-\epsilon_{k'})/\hbar} + \sum_k n_k\chi_j^{(k)}\chi_l^{(k)*} \tag{5.25}$$

with $N_\psi = \sum_q (1-n_q)|\chi_i^{(q)}|^2$. The simplest example is that of plane waves $\chi_j^{(k)} = \frac{1}{\sqrt{N_s}}e^{ikx_j}$ on a

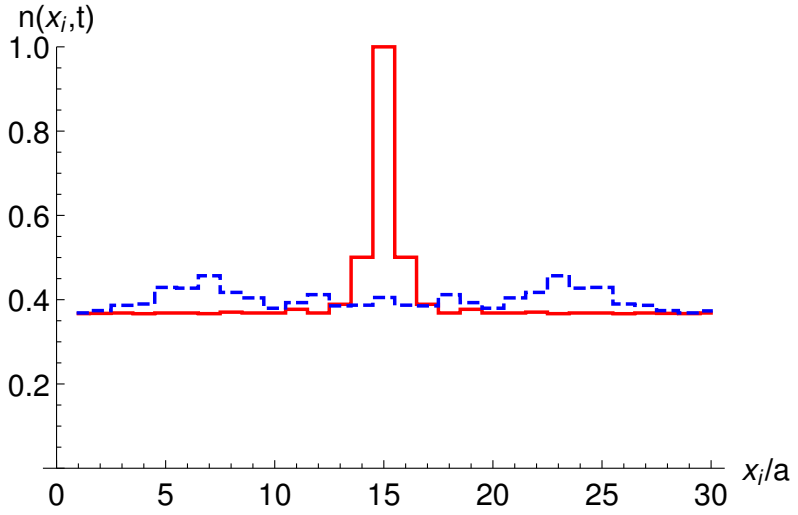


Figure 5.1: For a single band model on a 1D lattice with 30 sites, we assume the Fermi sea to consist of 11 electrons. At time $t_0 = 0$, we apply the electron creation operator of the center site. The resulting wave function has 12 electrons and must be renormalized, so the Pauli principle is effectively pushing existing electrons out of the way. The full line shows the electron density $n(x_i, t_0)$ immediately after the excitation. The dashed line shows the density at a later time, $n(x_i, t/\hbar = 10)$. The horizontal offset is due to the occupation by the Fermi sea, N_{occ}/N_s.

1D lattice with $\epsilon_k = 1-\cos(k)$, and zero temperature, $n_k = \theta(E_F - \epsilon_k)$. N_s is the number of sites, and N_{occ} the number of states occupied by the Fermi sea. Then $N_\psi = \frac{1}{N_s}\sum_l (1-n_l) = \frac{N_s - N_{occ}}{N_s}$ is simply the fraction of unoccupied states in $|FS\rangle$. The resulting time dependent local density for the plane wave example is

$$n(x_j, t) = -i\hbar G^<_{jj}(t,t) = \frac{1}{N_s(N_s - N_{occ})}\left|\sum_k \theta(\epsilon_k - E_F) e^{ik(x_j - x_i) + it\epsilon_k/\hbar}\right|^2 + \frac{N_{occ}}{N_s} \qquad (5.26)$$

where $x_j = ja$ is the coordinate of lattice site j. $n(x_j, t)$ is plotted in Fig. 5.1 immediately after the injection of the electron at the center coordinate (solid red line) and at a later time (dashed blue line). The Pauli principle does not always allow an electron to be put on the center site, since double occupation is forbidden for fermions. We had to renormalize the wave function $|\psi\rangle$ so that it describes a meaningful $(N_{occ}+1)$-electron state. This means that electrons at the site of injection are effectively pushed away, and the peak of $n(x_j, t_0)$ shows a broadening by the scale of the Fermi wave length. We have intentionally shown a discrete plot, since (5.26) makes only sense when evaluated at the discrete site coordinates $x_j = ja$. If the grid is refined, the peak becomes narrower because more electrons will be allowed on the lattice.

A similar interpretation would be possible for the hole correlation function $G^>$, in terms of holes (missing electrons). Let us emphasize that this section is only about the physical interpretation of $G^<$ (or $G^>$). Since the Green's functions themselves, and not many-particle states, will be the central objects for calculations, the results shown here do not impair the correctness of subsequent calculations in the context of a many particle system without interactions, if the Fermi sea is replaced by the vacuum state. In contrast to $G^<$ and $G^>$, in the non-interacting

system, G^R and G^A do not depend on the occupation of states and can directly be obtained from the Hamiltonian, as seen in (5.13), so their interpretation does not differ in a single- or many-particle picture.

5.4 The kinetic equations

Now, let us proceed to calculate first $G^{R/A}(E)$ and then $G^{</>}(E)$ of the non-equilibrium system. The inverse of a quadratic block matrix can be written in terms of its Schur complement $S_c = (A - BD^{-1}C)^{-1}$, keeping the block structure,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} S_c & -S_c B D^{-1} \\ -D^{-1} C S_c & D^{-1} + D^{-1} C S_c B D^{-1} \end{pmatrix}. \tag{5.27}$$

It is straightforward to check this very useful formula by matrix multiplication. We apply (5.27) to find the block structure of the retarded Green's function of the sample connected to a single lead p,

$$\begin{aligned} \begin{pmatrix} G_S^R & G_{Sp}^R \\ G_{pS}^R & G_p^R \end{pmatrix} &= \begin{pmatrix} E - H_S & -\tau_p \\ -\tau_p^\dagger & E + i\eta - H_p \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (E - H_S - \tau_p g_p^R \tau_p^\dagger)^{-1} & G_S^R \tau_p g_p^R \\ g_p^R \tau_p^\dagger G_S^R & g_p^R + g_p^R \tau_p^\dagger G_S^R \tau_p g_p^R \end{pmatrix}. \end{aligned} \tag{5.28}$$

Here and in the following, upper case symbols are used for the properties of the coupled system, like G for Green's functions. Lower case symbols like g are used for a decoupled system, which is assumed to be in equilibrium. We define the retarded and advanced self-energies of lead p as

$$\Sigma_p^R = \tau_p g_p^R \tau_p^\dagger, \quad \Sigma_p^A = \tau_p g_p^A \tau_p^\dagger = (\Sigma_p^R)^\dagger. \tag{5.29}$$

We see that in the inhomogeneous Schrödinger equation (5.9), we can substitute $H \to H_S + \Sigma_p^R$, in order to obtain the effective equation restricted to the sample region, but including the effect of lead p. We can apply (5.27) again to find the generalization to more leads. In the general case, we find

$$G_S^R = \left(E - H_S - \Sigma^R\right)^{-1} \quad \text{with } \Sigma^R = \sum_p \Sigma_p^R \tag{5.30}$$

and $G_S^A = (G_S^R)^\dagger$, $\Sigma^A = (\Sigma^R)^\dagger$. Note that the self-energy is energy dependent, so we cannot use it to reformulate the eigenproblem, i.e. the homogeneous Schrödinger equation.

The central results in the NEGF formalism are the kinetic equations

$$G^< = G^R \Sigma^< G^A \quad G^> = G^R \Sigma^> G^A \tag{5.31}$$

which relate $G^<$ to the inscattering self-energy $\Sigma^<$ and in a similar fashion, $G^>$ to the outscattering self-energy $\Sigma^>$, the latter being equivalent to the inscattering of holes. Here we give a simple derivation of (5.31) along the lines of [Dat07]. Consider the inhomogeneous Schrödinger equation (5.9) and an arbitrary excitation $S_j(t')$ at the site j of the sample. The response is given by $\psi_i(t) = \int dt' G_{ij}^R(t,t') S_j(t')$. We are interested in the correlation function $\psi_i(t)\psi_j^*(t')$ of the

response. As discussed in the previous section, the correspondence $\psi_i(t) \to c_i(t)$, $\psi_j^*(t') \to c_j^\dagger(t')$ can be used to find the correlation function

$$G_{ij}^<(t,t') = \frac{i}{\hbar}\langle c_j^\dagger(t')c_i(t)\rangle = \frac{i}{\hbar}\int dt''\int dt'''\sum_{n,m} G_{im}^R(t,t'')G_{nj}^A(t''',t')\langle S_n^\dagger(t''')S_m(t'')\rangle. \quad (5.32)$$

In formal analogy to $G^<$, we define the inscattering self-energy as correlation of the excitations in the sample region,

$$\Sigma_{ij}^<(t-t') = \frac{i}{\hbar}\langle S_j^\dagger(t')S_i(t)\rangle. \quad (5.33)$$

Here we also assume that the correlation depends only on $t-t'$, which makes sense if we consider that the excitations originate from scattering of the lead's electrons into the sample, and the combined system of leads and sample is assumed to be a steady state. Defining the energy representation of the inscattering function as usual, $\Sigma^<(E) = \frac{1}{\hbar}\int dt\, e^{iE(t-t')/\hbar}\Sigma^<(t-t')$, (5.32) becomes the matrix equation $G^<(E) = G^R(E)\Sigma^<(E)G^A(E)$. In a similar fashion, the inscattering functions of holes (or outscattering function of electrons) is defined as $\Sigma_{ij}^>(t-t') = -\frac{i}{\hbar}\langle S_i(t)S_j^\dagger(t')\rangle$. Remembering that $G^>$ denotes the correlation function for holes, one can prove the other kinetic equation, $G^> = G^R\Sigma^>G^A$.

Up to now, we have considered excitations in the sample region. However, we are rather interested in the sample's response to excitations in the leads, since those are the sources of the current, and their energy distribution is known. For this, we need the part G_{Sp}^R of the Green's function, see (5.28). Denoting by $S_{p,i}(t)$ an excitation in lead p, at site i, we can define an inscattering function of the isolated lead, $\sigma_{p,ij}^<(t-t') = \frac{i}{\hbar}\langle S_{p,j}^\dagger(t')S_{p,i}(t)\rangle$. Since $G_{Sp}^R = G_S^R\tau_p g_p^R$, the relation between the inscattering functions of sample and lead p is

$$\Sigma^< = \tau_p g_p^R \sigma_p^< g_p^A \tau_p^\dagger. \quad (5.34)$$

Next, we employ the kinetic equation (5.31) for the lead-p subspace,

$$G_S^< = G_S^R\tau_p(g_p^R\sigma_p^<g_p^A)\tau_p^\dagger G_S^A = G_S^R\tau_p g_p^<\tau_p^\dagger G_S^A. \quad (5.35)$$

Since the isolated lead (remember the small g stands for the uncoupled system) is in equilibrium, given by Fermi distribution function f_p and the spectral density $a_p = i(g_p^R - g_p^A)$, we can use (5.15) to find the lead-p contribution to $\Sigma^<$,

$$\Sigma_p^< = \tau_p g_p^<\tau_p^\dagger = i f_p\tau_p a_p\tau^\dagger =: i f_p\Gamma_p. \quad (5.36)$$

Further, we have defined the broadening[1] matrices

$$\Gamma_p = i(\Sigma_p^R - \Sigma_p^A). \quad (5.37)$$

Γ_p is related to the velocity operator of lead p, along the direction to which it extends, as we will see later, by the orthogonality relation of modes (5.82). By analogy, the outscattering function of the lead is given by the equilibrium distribution function for holes,

$$\Sigma_p^> = \tau_p g_p^>\tau_p^\dagger = -i(1-f_p)\Gamma_p. \quad (5.38)$$

[1] Datta [Dat05] uses the terminology *broadening matrix* for Γ_p, because the coupling to the leads is responsible for smoothing out discrete levels in the density of states of a finite sample, which results in a continuous density of states for the coupled system.

Note that in our derivation, we have not used the fact that by changing $g_p^R \to g_p^<$ in the formula (5.29) for Σ_p^R, we obtain $\Sigma_p^<$. This feature is nice but currently seems mysterious if we remember, that the physical meaning of Σ_p^R is completely different from that of $\Sigma_p^<$. However, if one derives the expression for the self-energy in Keldysh space, this feature is natural since both Σ_p^R and $\Sigma_p^<$ can be obtained from the same Keldysh self-energy. In general, the Langreth rules (as first used in [Lan+72]) can be used to find real-time representations of products in Keldysh space. From these rules, it can be understood that it is indeed possible to interchange the superscripts $<$ and R, if they appear only once on each side of the equation.

5.5 Calculating the currents

Now that the stage has been set by introducing the most important definitions and equations, we want to apply this to calculate the currents leaving or entering the sample from the leads. There are two ways to do this, and we will discuss them both in order to have a versatile toolbox for calculations and interpretations. Firstly, we can find a current operator representation in the lattice model which expresses the current density in terms of $G^<$, and then integrate the current leaving the sample in direction of some chosen lead. This means we will have to integrate over the surface separating sample and lead regions, and enter $G^<$ as given by the kinetic equation. This way has the advantage that the connection to the local current density is clear. The other possibility is to consider eigenmodes of the leads, and derive a scattering matrix, called S-matrix, in the basis of these modes. The current will be proportional to a sum of scattering probabilities. The latter approach corresponds to using a mode basis for evaluating the integral over the surface between sample and lead. The advantage is that the scattering matrix is a very useful and simple tool to analyse the transport properties and can also be used to combine several mesoscopic devices in a phase coherent way.

In order to understand where the current flows, we have to further split up the sample in different regions. We will denote the sample region by S, and the lead regions by p and q (again we just discuss the case of two leads, the generalization being immediate). The lattice representation of the Hamiltonian will always (in our cases) take the form of a tight binding model, i.e. it contains only coupling terms to neighbour and sometimes next-neighbour sites. Therefore, we can define a region $p' \subset S$ located next to lead p, containing all sites with non-zero couplings to lead p, and likewise $q' \subset S$. We assume that they do not intersect and that there is no direct coupling between p' and q'. This assumption is no restriction, since it can always be met by increasing the sample region at cost of the leads, which will not change the lead currents. By s (lower case), we denote the remaining region in S, i.e $s = S \setminus (p' \cup q')$. All these regions correspond to index sets for lattice sites. So e.g. in $\tau_{p,ij} \neq 0$, the index $i \in p'$ and $j \in p$. This means, while have already demanded earlier that $\hat{\tau}_q \hat{\tau}_p^\dagger = 0$, now we also demand that $\hat{\tau}_q^\dagger \hat{\tau}_p = 0$. We can partition H_S in blocks analogue to (5.1),

$$H_S = \begin{pmatrix} H_s & H_{sp'} & H_{sq'} \\ H_{sp'}^\dagger & H_{p'} & \\ H_{sq'}^\dagger & & H_{q'} \end{pmatrix}. \tag{5.39}$$

For the total charge in region s, we introduce the operator $\hat{Q}(t) = -e \sum_{k \in s} c_k^\dagger(t) c_k(t)$. The

continuity equation $\dot{Q}(t) = -\int_s \mathrm{d}V\, \nabla \cdot \mathbf{J}(t)$ tells us that we can define currents by considering

$$\begin{aligned}\dot{Q}(t) &= \frac{-ie}{\hbar}\langle[\hat{H}, \hat{Q}(t)]\rangle = \frac{-ie}{\hbar}\sum_{k\in s}\sum_{i\in S}\left(H_{ik}\langle c_i^\dagger(t)c_k(t)\rangle - H_{ki}\langle c_k^\dagger(t)c_i(t)\rangle\right) \\ &= -e\mathrm{Tr}_s[G^<(t,t), H],\end{aligned} \tag{5.40}$$

where Tr_s denotes the trace over sites belonging to region s. We use the same partitioning of $G_S^<$ as for H_S. Writing out matrix products in space S in the blocks and making use of the cyclic invariance of the trace over space s, we arrive at

$$\mathrm{Tr}_s[G^<(t,t), H] = \sum_{p'} \mathrm{Tr}_s\left[G^<_{sp'}(t,t)H_{p's} - H_{sp'}G^<_{p's}(t,t)\right]. \tag{5.41}$$

Now we see how $\int_s \mathrm{d}V\, \nabla \cdot \mathbf{J}(t)$ divides up into contributions from each lead p, and we also see that it only depends on the surface terms, analogue to Gauss' theorem $\int_s \mathrm{d}V\, \nabla \cdot \mathbf{J} = \int_{\partial s} \mathrm{d}\mathbf{S} \cdot \mathbf{J}$. Since we assume a steady state, we can write $G^<(t,t) = \frac{1}{2\pi\hbar}\int \mathrm{d}E\, G^<(E)$. In the steady state, the total current going out of the sample region is zero, but the currents through individual leads can be non-zero. We identify the current flowing into lead p as $I_p = \int \mathrm{d}E\, I_p(E)$ with

$$I_p(E) = \frac{e}{h}\mathrm{Tr}_s\left[G^<_{sp'}(E)H_{p's} - H_{sp'}G^<_{p's}(E)\right]. \tag{5.42}$$

Next, we introduce a projector $P_{p'}$ projecting onto the sites of the region p'. This will allow us to rewrite the lead-p current in terms of matrices defined on the sample space S. Using the cyclic invariance under Tr_s and $\mathrm{Tr}_{p'}$, we can add up a zero term and extend the trace and all appearing matrices to the larger region S,

$$\begin{aligned}I_p(E) =& \frac{e}{h}\mathrm{Tr}_{p'}\left[H_{p's}G^<_{sp'} - G^<_{p's}H_{sp'}\right] + \frac{e}{h}\mathrm{Tr}_{p'}\left[H_{p'p'}G^<_{p'p'} - G^<_{p'p'}H_{p'p'}\right] \\ =& \frac{e}{h}\mathrm{Tr}_S P_{p'}\left[H_S G_S^< - G_S^< H_S\right].\end{aligned} \tag{5.43}$$

Using the definitions (5.30) of G_S^R, G_S^A and the kinetic equation (5.31), we find

$$H_S\, G_S^< - G_S^< H_S = H_S\, G_S^R\, \Sigma^< G_S^A - G_S^R\, \Sigma^< G_S^A H_S \tag{5.44}$$

$$H_S\, G_S^R = E G_S^R - \Sigma^R\, G_S^R - \mathbf{1} \tag{5.45}$$

$$G_S^A\, H_S = E G_S^A - G_S^A \Sigma^A - \mathbf{1}, \tag{5.46}$$

so

$$H_S\, G_S^< - G_S^< H_S = G_S^R\, \Sigma^< - \Sigma^<\, G_S^A - \Sigma^R\, G_S^< + G_S^<\, \Sigma^A. \tag{5.47}$$

In the above expressions, the self-energies $\Sigma^{R/A}$ and inscattering functions $\Sigma^<$ are still the sum over all leads. Using the cyclic invariance of the trace and $P_{p'}\hat{\tau}_p = \hat{\tau}_p$, $P_{p'}\hat{\tau}_q = 0$, we find

$$P_{p'}\Sigma^{R/A/<} = \Sigma_p^{R/A/<}, \quad \Sigma^{R/A/<}P_{p'} = \Sigma_p^{R/A/<} \tag{5.48}$$

and finally

$$I_p(E) = \frac{e}{h}\mathrm{Tr}\left[\Sigma_p^<(G_S^R - G_S^A) - (\Sigma_p^R - \Sigma_p^A)G_S^<\right] = \frac{e}{ih}\mathrm{Tr}\left[\Sigma_p^< A_S - \Gamma_p\, G_S^<\right]. \tag{5.49}$$

As we see, the expressions for the current at different leads depend on the same $G_S^<$ and spectral density A_S of the sample region. $\Sigma_p^<$ is responsible for the inscattering of electrons from lead p into the sample region and depends on the chemical potential μ_p by the Fermi distribution. The inscattering further depends on the available states of the sample, as described by A_S. On the other hand, the outscattering term $G_S^<\Gamma_p$ depends on the occupation of states in the sample, given by $G_S^<$, and the available states in the lead p, which are described by the spectral density $a_p = i(g_p^R - g_p^A)$ of the lead p in $\Gamma_p = \tau_p a_p \tau_p^\dagger$.

Now we can specialize to inscattering and outscattering given by the leads. For this we combine the expression for spectral density (5.12) with the kinetic equations (5.31) and the equilibrium inscattering function $\Sigma_p^< = i f_p \Gamma_p$ of (5.36), to find the spectral density

$$A_S = i(G_S^> - G_S^<) = \sum_q G_S^R \Gamma_q G_S^A. \tag{5.50}$$

Similarly, we use the kinetic equation and the equilibrium outscattering function (5.38), and find the Landauer-Büttiker formula for the current at lead p

$$I_p = \frac{e}{h} \sum_{q \neq p} \int dE\ T_{pq}(E)(f_p(E) - f_q(E)). \tag{5.51}$$

The current is given by lead-dependent Fermi distributions $f_p = (e^{(E-\mu_p)/k_B T_p} + 1)^{-1}$, and transmission probabilities T_{pq} for scattering from lead q to lead p, counting contributions from any modes of the leads (since the T_{pq} are sums of probabilities for independent modes, they may be greater than unity). They are defined as the traces

$$T_{pq} = \mathrm{Tr}[\Gamma_p G^R \Gamma_q G^A] \quad (p \neq q). \tag{5.52}$$

Only for $p \neq q$, these traces should be understood as transmission probabilities. If needed, reflection probabilities $T_{pp} = N_p - \sum_{q \neq p} T_{pq}$ may be obtained from current conservation, which results in a sum rule for the T_{pq}, see Eq. (5.73).

If all leads are at the same temperature T, but a small electrical bias is applied so that the potentials μ_p differ, we can expand f_p and f_q around a common potential, which will be the Fermi energy E_F. Using the notation $f_0(E) = 1/(e^{E/k_B T} + 1)$ and assuming $|\mu_p - \mu_q| \ll k_B T$ for the expansion of Fermi functions, we find

$$I_p = \frac{e}{h} \sum_{q \neq p} \int dE\ T_{pq}(E) \frac{-\partial f_0(E - E_F)}{\partial E} (\mu_p - \mu_q). \tag{5.53}$$

In the zero temperature limit, the Fermi functions become step functions $f_p = \theta(E_F - \mu_p)$ and (5.53) reduces to

$$I_p = \frac{e}{h} \sum_q T_{pq}(E_F)(\mu_p - \mu_q) = \frac{e}{h} \sum_q T_{qp}(E_F)\mu_p - T_{pq}(E_F)\mu_q \tag{5.54}$$

where we have made use of the sum rule (5.73) in the last step. This formula allows for a simple interpretation of the current. The contribution $T_{qp}(E_F)\mu_p$ is due to the current flowing from lead p to other leads, while the contributions $T_{pq}(E_F)\mu_q$ are currents coming in from all other leads. Since μ_p/e are electrical voltages, we recognize that the quantum of conductance is $\frac{e^2}{h}$,

not including a factor two for spin degeneracy. Quantization of the conductance in steps of $\frac{2e^2}{h}$ has been experimentally observed [Wee+88].

Interestingly, the formula (5.53) is even valid for finite bias, i.e. when the condition $|\mu_p - \mu_q| \ll k_B T$ is not met at low temperatures, as long as the transmission functions T_{pq} are constant in the energy range $[\mu_p, \mu_q]$. By rewriting (5.51) in the form of a convolution, where the convolution kernel is related to finite temperature and causes smoothing of the function $T_{pq}(E)$ on the scale of $k_B T$, it becomes clear that the allowed bias is $|\mu_p - \mu_q| \leq \max(k_B T, \delta)$, if the energy δ is the scale on which $T_{pq}(E)$ fluctuates appreciably [Dat07]. In clean (ballistic) samples, δ can be large, while in disordered samples, universal conductance fluctuations will give a highly oscillating signal $T_{pq}(E)$ at low temperatures. However, in a situation of finite bias, the transmission function T_{pq} can also change due to the electrical field. To capture this effect in the formalism, a self-consistent solution of the Poisson equation for the charge distribution would be required. The idea of rewriting the integral (5.51) in the form of a convolution is also used in Chapter 9.2.3, where we derive a Mott-like relation. It relates charge conductance to transverse spin currents that are induced by a temperature gradient, and it is valid at finite temperature.

Since the transmission function T_{pq} essentially describes the mesoscopic conductance matrix which relates the vector of lead currents to the vector of lead voltages, it is interesting to compare with the Kubo formula for the conductivity tensor (compare Appendix A). The matrices Γ_p, Γ_q in formula (5.52) can be seen as current operators for the leads, showing the similarity to the Kubo conductivity (A.7), which also involves the correlator of two current operators. The Kubo formula is an application of the dissipation-fluctuation theorem, because it relates equilibrium fluctuations to a dissipative conductivity. Since the transmission formula (5.52) is the result for a phase coherent mesoscopic sample and dissipation enters only by the coupling to the leads, averaging over an ensemble of uncorrelated disorder configurations is needed to effectively introduce decoherence and meaningfully compare with the Kubo formula [Dat07]. In a phenomenological way, it is also possible to introduce decoherence by adding artifical floating leads to the sample, which have no net effect on the current, but destroy the phase relation entering and leaving electrons.

All of the discussion until here has been for charge or particle currents. However in this thesis, we will be especially interested in spin currents, i.e. in the transport of the angular momentum $\frac{\hbar}{2}$ of the electron spin. In a system where $S_z = \frac{\hbar}{2}\sigma_z$ is conserved, the current of the z component of spin can be defined as $\frac{\hbar}{2}(I_\uparrow - I_\downarrow)$, where $I_{\uparrow/\downarrow}$ are particle currents of electrons in eigenstates of S_z. It is common to define the local spin density $n^{(\mu)}(\mathbf{r})$ and the local spin current $J_i^{(\mu)}(\mathbf{r})$ of the vector component μ of spin, in a state $|\psi\rangle$ of the system, as

$$n^{(\mu)}(\mathbf{r}) = \frac{\hbar}{2}\psi^\dagger(\mathbf{r})\sigma_\mu\psi(\mathbf{r}) = \langle\psi|\frac{\hbar}{2}\delta(\mathbf{r}-\hat{\mathbf{r}})\sigma_\mu|\psi\rangle \equiv \langle\psi|\hat{n}^{(\mu)}(\mathbf{r})|\psi\rangle, \tag{5.55}$$

$$J_l^{(\mu)}(\mathbf{r}) = \frac{\hbar}{2}\psi^\dagger(\mathbf{r})\{\sigma_\mu, \frac{\partial H}{\hbar\,\partial k_l}\}\psi(\mathbf{r}) = \langle\psi|\frac{1}{4}\left\{\delta(\mathbf{r}-\hat{\mathbf{r}}), \{\sigma_\mu, \frac{\partial H}{\partial k_l}\}\right\}|\psi\rangle \equiv \langle\psi|\hat{J}_l^{(\mu)}(\mathbf{r})|\psi\rangle. \tag{5.56}$$

If we define $\sigma_0 = 1$, choosing $\mu = 0$ gives the usual charge density and charge current up to a prefactor $\frac{2e}{\hbar}$. On the right side of (5.55) and (5.56), we have defined corresponding operators for the local spin density and local spin currents.

For numerical treatment, lattice representations of these operators are needed. They can be found by evaluation of matrix elements within a basis of states $|\mathbf{r}_i, s\rangle$ which are localized on lattice sites $\mathbf{r}_i$, and include as a factor a spin-z basis state $|s\rangle \in \{|\uparrow\rangle, |\downarrow\rangle\}$. Further, a finite difference method needs to be applied to the Hamiltonian H to find matrix elements $H_{\mathbf{r}_i s, \mathbf{r}_j s'}$,

see Equations (2.33), (2.34) and (2.35). We find lattice representations corresponding to the operators in (5.55) and (5.56)

$$\langle \mathbf{r}_i, s|\hat{n}^{(\mu)}(\mathbf{r})|\mathbf{r}_j, s'\rangle = \frac{\hbar}{2}\delta_{\mathbf{r},\mathbf{r}_i}\delta_{\mathbf{r}_i,\mathbf{r}_j}(\sigma_\mu)_{ss'}, \tag{5.57}$$

$$\langle \mathbf{r}_i, s|\hat{\mathbf{J}}^{(\mu)}(\mathbf{r})|\mathbf{r}_j, s'\rangle = \frac{1}{4i}(\delta_{\mathbf{r},\mathbf{r}_i} + \delta_{\mathbf{r},\mathbf{r}_j})(\mathbf{r}_i - \mathbf{r}_j)\{\sigma_\mu, H\}_{\mathbf{r}_i s,\mathbf{r}_j s'}, \tag{5.58}$$

$$\{\sigma_\mu, H\}_{\mathbf{r}_i s,\mathbf{r}_j s'} = \sum_{s''}\left((\sigma_\mu)_{ss''}H_{\mathbf{r}_i s'',\mathbf{r}_j s'} + H_{\mathbf{r}_i s,\mathbf{r}_j s''}(\sigma_\mu)_{s''s}\right). \tag{5.59}$$

If the Hamiltonian involves only linear and quadratic terms in $\hat{\mathbf{k}}$, its lattice representation contains only hopping elements connecting nearest neighbour sites, which greatly reduces the number of matrix elements needed in (5.58). Lattice models including only nearest or next-nearest neighbour couplings are also called tight binding models, and their matrix representations are sparse matrices. For numerical treatment, only the non-zero entries of sparse matrices need to be stored in memory. We use the efficient linear solver MUMPS [Ame+01] (freely available under `http://mumps.enseeiht.fr`), which works with sparse matrices and can benefit from parallelization.

The observables can be expressed in terms of the lesser Green's function (5.4), in the way

$$\mathbf{J}^{(\mu)}(\mathbf{r}) = \langle\hat{\mathbf{J}}^{(\mu)}(\mathbf{r})\rangle = \sum_{\mathbf{r}_i,\mathbf{r}_j,s,s'}\langle \mathbf{r}_i, s|\hat{\mathbf{J}}^{(\mu)}(\mathbf{r})|\mathbf{r}_j, s'\rangle\int dE\,\frac{-i}{2\pi}G^{<}_{\mathbf{r}_j s',\mathbf{r}_i s}(E) \tag{5.60}$$

in units of spin current per lattice site, and analogue for the densities (charge or spin per lattice site). This allows for numerical evaluation, based on the kinetic equations (5.31) and assumed inscattering from the leads, i.e. for some choice of lead potentials and temperatures. In Chapter 9.4 the non-equilibrium part of the spin current induced by an electrical bias is visualized in this way, and in Chapter 10.4, we define an helicity operator which is associated with spin and plot the corresponding bias-induced density.

Being interested in systems that generate spin accumulation or spin currents, we will need spin-orbit coupling terms in the Hamiltonian. (Other possibilities would be to start with a system where spins $\uparrow,\downarrow$ are populated differently already at equilibrium, like in a ferromagnet, or to induce a non-equilibrium spin population optically, but we will not be interested in such approaches.) In the presence of SO coupling, the spin current will not be conserved. As shown in Appendix E, it obeys the generalized continuity equation

$$\frac{\partial}{\partial t}n^{(\mu)} + \nabla\cdot\mathbf{J}^{(\mu)} = \mathcal{T}_\mu, \tag{5.61}$$

which involves a source term given by the torque $\mathcal{T}_\mu = \frac{\hbar}{2}\mathrm{Re}\left(\psi^\dagger(\mathbf{r})\frac{1}{i\hbar}[\sigma_\mu, H]\psi(\mathbf{r})\right)$. Clearly, this is a problem if a well-defined spin current should be measured. Moreover, in a time reversal but not spatial inversion symmetric system, the spin current (5.56) can be non-zero even at thermodynamic equilibrium [Ras03]. On the other hand, there is no problem with the definition of spin densities, and non-zero spin accumulation at equilibrium is not possible in a time reversal symmetric system. There have been attempts to solve the problem of non-conservation of spin currents in different ways. One proposal has been to incorporate the source term in the definition of spin current, which effectively corresponds to considering the continuity equation of the operator $S_\mu\hat{\mathbf{r}}$, instead of S_μ [Shi+06]. This has the advantage that Onsager reciprocal

relations, i.e. symmetric linear relations between a vector of thermodynamic forces like electrical fields, temperature gradients and pressure, and a vector of the corresponding charge, heat and particle currents, can be extended to include spin currents. For some discussion including other definitions of spin currents, see [Nik+09]. The easiest solution to the problem of non-conservation is to exclude SO coupling from those parts of the sample where the spin current should be measured, e.g. in some of the leads [Nik+05a; Han+04]. Then, the lead Hamiltonian is diagonal in spin space and separate leads may be defined for spin-↑ and spin-↓, to obtain a straightforward spin-dependent generalization of the NEGF formalism. This approach is used in Chapters 7.3 and 9. For this approach, one must decide on the spin vector component that one wants to analyse.

If one wishes to analyse different spin vector components within a single calculus, or if the transport of a non-conserved quantity should be analysed in terms of contributions carried by individual leads, one can combine transmission amplitudes as given by the scattering matrix (S-matrix) with operator expectation values. Even if the spin current oscillates in a lead, its average may constitute a non-zero and meaningful physical quantity, if the torque term in (5.61) does not cancel the signal. In the following Section 5.6 we will focus on the scattering formalism in terms of modes of a lead. Since the S-matrix provides phase relations between different modes of a lead, spin currents of different directions may be found from it. This is discussed in more detail in Appendix D and will be applied in Chapter 10. We just give the main result here, which is rather intuitive. The expectation value of a local operator $\hat{O}(\mathbf{r})$, averaged over lead p, in response to a bias $\delta\mu_q$ applied at a different lead q, and evaluated at $T = 0$, is given by

$$\lim_{V\to\infty} \frac{1}{V} \int_V d\mathbf{r}^2 \frac{\delta\langle\hat{O}(\mathbf{r})\rangle}{\delta\mu_q} = \frac{1}{\hbar} \sum_{m,n,l}^{v_m=v_n} t_{m,l} t^*_{n,l} \frac{1}{|v_n|} O_{n,m} \tag{5.62}$$

where V is the area of averaging, m, n, l are modes of lead q and the summation is restricted to pairs with mode velocities $v_m = v_n$. $t_{m,l}$ are entries of the S-matrix and $O_{n,m}$ are matrix elements of $\hat{O}(\mathbf{r})$ with respect to the lead's eigenmodes. The result is the same as when $\langle\hat{O}(\mathbf{r})\rangle$ is evaluated with the kinetic equation (5.31) and averaged over the (semi-infinite) lead.

5.6 Transverse modes of the leads and the scattering matrix

The eigenmodes of the lead l are needed to write down transmission amplitudes in the form of a Fisher-Lee relation, and are also needed to construct the lead's Green's function g^R_l and self-energy Σ^R_l. We choose a lead-dependent local coordinate system, where $x_l \to \infty$ corresponds to the direction of the infinite extension of the lead, and we define basis vectors $|j\rangle$ to be localized on sites with coordinate $x_l = ja$, a being the lattice constant. The isolated lead l is assumed to be invariant under translations by the lattice constant a in x_l-direction, and we can write its Hamiltonian as

$$H_l = \sum_{j=0}^{\infty} H_0|j\rangle\langle j| + H_1|j\rangle\langle j+1| + H_{-1}|j+1\rangle\langle j| \tag{5.63}$$

where H_0, H_1 and $H_{-1} = H_1^\dagger$ are $N \times N$ matrices, with $N = N_s N_b$ for a lead with N_s sites per slice in transverse direction (y_l-coordinate) and a Hamiltonian including N_b bands. H_1 is the coupling term between neighbouring slices, and H_0 the Hamiltonian of an isolated slice. Usually,

these matrices are obtained from the discretization of an effective multiband Hamiltonian which does not distinguish between sample and leads, compare Chapter 2.4.

To develop the scattering formalism, it is easier to start from the corresponding infinite lead, which extends from $x_l = -\infty$ to ∞,

$$H_{l\infty} = \sum_{j=-\infty}^{\infty} H_0|j\rangle\langle j| + H_1|j\rangle\langle j+1| + H_{-1}|j+1\rangle\langle j| \tag{5.64}$$

We make the plane wave ansatz $\langle x_l|\phi_{n,k}\rangle = ce^{ikx_l}\phi_{n,k}$, where the transverse mode vectors are normalized to unity, $|\phi_{n,k}|^2 = 1$, and the normalization constant is chosen as $c = |v_n|^{-\frac{1}{2}}$ for propagating modes with velocity v_n. With this convention, current conservation will imply that the transmission coefficients are entries of a unitary matrix. For evanescent states, i.e. when $\mathrm{Im}(k) \neq 0$, the normalization constant c is unimportant. We obtain the effective infinite-lead Hamiltonian

$$H_{l\infty}(k)\phi_{n,k} = \left(e^{ika}H_1 + H_0 + e^{-ika}H_{-1}\right)\phi_{n,k} = E_n(k)\phi_{n,k}. \tag{5.65}$$

The solutions are the transverse eigenmodes $\phi_{n,k}$, and for fixed k, they provide an orthonormal basis. However, instead of the infinite set of solutions, we only need the finite set at a fixed energy E, to analyse the transport at this energy. Writing $\lambda = e^{ika}$, (5.65) becomes a quadratic eigenvalue problem for λ,

$$\left((E - H_0)\lambda - H_1\lambda^2 - H_{-1}\right)\phi = 0 \tag{5.66}$$

of size $N \times N$. For the numerical treatment, we map it to a linear eigenvalue problem of size $2N \times 2N$, for the vector $(\phi, \lambda\phi)$,

$$\underbrace{\begin{pmatrix} 0 & \mathbb{1} \\ -H_1^{-1}H_{-1} & H_1^{-1}(E - H_0) \end{pmatrix}}_{=:W} \begin{pmatrix} \phi \\ \lambda\phi \end{pmatrix} = \lambda \begin{pmatrix} \phi \\ \lambda\phi \end{pmatrix}. \tag{5.67}$$

In our applications of the formalism, H_1 will always be invertible, but note that this is not always the case, and the formalism may be modified for such situations [Wim08; Run+08]. We can use standard numerical routines to calculate the set of solutions $\{\lambda_n = e^{ik_n a}\}_n$ and $\{\phi_n\}_n$ at fixed energy E. Note that the ϕ_n are in general no longer orthogonal, since they correspond to different k_n. Only in the simplest case, when there is no magnetic field and no SO coupling, the ϕ_n are independent of k and thus orthogonal.

There is a total of $4N$ solutions λ_n, and they fall into four groups. First, there are propagating solutions, for which $|\lambda_n| = 1$. We call a solution right-propagating, if the velocity $v_n = \frac{\partial E_n(k)}{\hbar\partial k}\Big|_{k=k_n} > 0$, and left-propagating, if $v_n < 0$. There is an equal number of right- and left-propagating solutions, since $E_n(k)$ is a periodic function, and each subband $E_n(k)$ must cross the energy E equally often with a positive slope ($v_n > 0$) and negative slope. Further, there are evanescent solutions with $|\lambda_n| \neq 1$, or equivalently $\mathrm{Im}\, k_n \neq 0$. They can be either right-decaying ($\mathrm{Im}\, k_n < 0$), or left-decaying ($\mathrm{Im}\, k_n > 0$). We can rewrite the determinant of the quadratic eigenvalue problem (5.66) in terms of its Hermitian conjugate,

$$0 = \frac{1}{\lambda^{*2N}} \det\left((E - H_0)\lambda - H_1\lambda^2 - H_{-1}\right) = \det\left((E - H_0)\frac{1}{\lambda^*} - H_{-1}\frac{1}{\lambda^{*2}} - H_1\right). \tag{5.68}$$

It follows that for any eigenvalue λ, $1/\lambda^*$ is also an eigenvalue. Therefore, for each right-decaying mode there is also a corresponding left-decaying mode. Like Wimmer [Wim08], we will call a mode right-going if it is right-propagating or right-decaying. For a lead that is attached at the right side of the sample, we will call right-going modes outgoing, since they propagate away from the sample, and ingoing, if the lead is attached at the left side of the sample. Likewise, we introduce the group of left-going modes, which are left-propagating or left-decaying. There are N left-going and N right-going, or equivalently N ingoing and N outgoing modes. If the nature of the mode is important, we denote them with with and index $>$ for ingoing and $<$ for outgoing, like $\phi_{n,>}$, $k_{n,>}$, $\lambda_{n,>}$, $v_{n,>}$ and analogue with $<$.

We still need a numerical prescription how to calculate the mode velocities,

$$v_n = \left.\frac{\partial E_n(k)}{\hbar\,\partial k}\right|_{k_n} = \phi_n^\dagger \frac{\partial H_{l\infty}(k_n)}{\hbar\,\partial k_n}\phi_n = \phi_n^\dagger \frac{ia}{\hbar}\left(\lambda_n H_1 - \lambda_n^{-1} H_{-1}\right)\phi_n. \tag{5.69}$$

In order to develop a scattering formalism and define transmission amplitudes, we need scattering states of the combined system of sample and leads. In the asymptotic region ($x_l \to \infty$), all evanescent contributions to the scattering waves have decayed. Without loss of generality, we can assume that a large part of the lead's area is included in sample, and therefore, we may restrict ourselves to the propagating modes, when expanding the scattering states in the modes of a lead. We make the ansatz for a scattering state corresponding to a wave which enters only through lead l and leaves by leads l'

$$\psi^{(l)}_{n,>}(x) = \begin{cases} |v_{n,>}|^{-\frac{1}{2}} e^{ik_{n,>}x_l}\phi^{(l)}_{n,>} + \sum_m t_{lm,ln}|v_{m,<}|^{-\frac{1}{2}} e^{ik_{m,<}x_l}\phi^{(l)}_{m,<} & x \in l \\ \sum_m t_{l'm,ln}|v_{m,<}|^{-\frac{1}{2}} e^{ik_{m,<}x_{l'}}\phi^{(l')}_{m,<} & x \in l' \neq l \end{cases} \tag{5.70}$$

and we do not care about the value of $\psi^{(l)}_{n,>}(x)$ in the sample region. Here, we have added a superscript (l) to identify the lead of the modes. x_l and $x_{l'}$ are to be understood as x, in the lead's local coordinate systems. $t_{lm,ln}$ are the reflection amplitudes for parts leaving by the same lead and $t_{l'm,l,n}$ are transmission amplitudes from lead l, mode n to lead l', mode m. These amplitudes together make up the scattering matrix (S-matrix).

We can show the unitarity of the S-matrix in the following way. We define a vector $\mathbf{a} = (a_{ln})_{ln}$, such that $|a_{ln}|^2$ is the current carried by the incoming mode n of lead l, and further define a vector $\mathbf{b} = S\mathbf{a}$ for the outgoing currents of all leads. Then current conservation implies

$$I_{\text{in}} = \mathbf{a}^\dagger\mathbf{a} = I_{\text{out}} = \mathbf{b}^\dagger\mathbf{b} = \mathbf{a}^\dagger S^\dagger S\mathbf{a} \quad \Rightarrow S^\dagger S = 1. \tag{5.71}$$

Here it is important to use the normalization constant $c = |v_n|^{-\frac{1}{2}}$ of propagating states. We further define transmission probabilities by $T_{pn,qm} = |t_{pn,qm}|^2$, and the total transmission from lead q to lead p by the sum over all contributing modes $T_{pq} = \sum_{n,m}|t_{pm,qn}|^2$. From the unitarity of S, we deduce the sum rules

$$\sum_{p,n} T_{pn,qm} = \sum_{q,m} T_{pn,qm} = 1, \tag{5.72}$$

$$\sum_q T_{pq} = \sum_q T_{qp} = N_p(E), \tag{5.73}$$

where $N_p(E)$ is the number of propagating modes in lead p for energy E.

In the numerical code, we do not calculate the scattering states (although we do calculate the modes needed for the asymptotic states), but rather calculate the retarded Green's function of the sample, which provides the same information. The relation between the transmission amplitudes and the Green's function is given by the Fisher-Lee Relation. In a simple form it was originally shown by Fisher and Lee [Fis+81], and in a general form which is valid also for non-orthogonal modes, it has been derived by Baranger and Stone [Bar+89] and by Wimmer, where the latter is particularly useful for us, since it is formulated for a lattice model. The general Fisher-Lee relation for the full S-matrix $S = (t_{pm,qn})_{pm,qn}$ including transmission and reflection coefficients is [Wim08]

$$t_{pm,qn} = \frac{1}{\sqrt{|v_{m,<} v_{n,>}|}} \phi^{(p)\dagger}_{m,<} \left(i\Gamma_p G^R \Gamma_q - \delta_{pq}\Gamma_p \right) \phi^{(q)}_{n,>}. \tag{5.74}$$

The matrix product notation $\Gamma_p G^R \Gamma_q$ implies that of G^R, only the submatrix corresponding to the sites adjacent to the leads p and q is needed, i.e. the surface Green's function of the sample. Compared the formula (5.52) for transmission probabilities T_{pq}, the full S-matrix (5.74) provides more information, since the distribution of the current among the modes and phase relations between modes can be analysed. Phase relations between modes can be important if S-matrices of two devices are combined to describe one large phase-coherent device (like it is done in Chapter 10.5). In this case, one has to keep track of the phase convention used for definition of the asymptotic states. Of course, phases can only be defined for propagation over finite distances. The two devices that are combined in Chapter 10.5 have a finite distance, so S-matrix phases are important in that case. On the other hand, if the S-matrix is used for evaluation of currents at an infinite distance from the sample, the complex phases of the $t_{pm,nq}$ are undefined.

The physics behind the Fisher-Lee relation is clear. One can take the an asymptotic state in lead q, propagate it by the Green's function G^R of the sample, and calculate the overlap with an asymptotic state in lead p, in order to obtain transmission amplitudes. The $\Gamma_{p/q}$ matrices play the role of lead-velocity operators, which provide an orthogonality relation for the transverse modes. A detailed derivation can be found in [Wim08], and here we give a simplified and less formal derivation. As in [Bar+89], we define a Hermitian operator counting the particle current passing from slice j to slice $j+1$ of a lead,

$$K(j) = \frac{ia}{\hbar} \left(H_1 |j\rangle\langle j+1| - H_{-1} |j+1\rangle\langle j| \right). \tag{5.75}$$

We can convince ourselves of this interpretation by a simple calculation. We introduce the operator projecting on the slice j of the lead, $P_j = |j\rangle\langle j|$. With the representation $\hat{x}_l = \sum_j jaP_j$, it is easy to check that the velocity operator is

$$\frac{1}{i\hbar}[\hat{x}_l, H_{l\infty}] = \sum_j K(j). \tag{5.76}$$

Then we find $K(j)$ by projecting the current operator onto the two relevant slices,

$$K(j) = (P_j + P_{j+1}) \left(\sum_i K(i) \right) (P_j + P_{j+1}). \tag{5.77}$$

Using the notation $\langle j|\phi_m\rangle = e^{ik_m ja}\phi_m$, we find velocity matrix elements for any two modes, regardless if propagating or evanescent,

$$\langle\phi_n|K(j)|\phi_m\rangle = \underbrace{\phi_n^\dagger \frac{ia}{\hbar}\left(H_1 e^{ik_m a} - H_{-1}e^{-ik_n a}\right)\phi_m}_{v_{nm}} e^{i(k_m-k_n)ja} = v_n\delta_{nm} \tag{5.78}$$

where the last equality follows from current conservation, which demands that (5.78) must not depend on j. Explicitly, this is seen as follows. Let us assume non-degenerate subbands $E_n(k)$, and consider a superposition of modes of the same energy, $|\psi\rangle = a|\phi_n\rangle + b|\phi_{n'}\rangle$. Then, due to current conservation, the current expression

$$\langle\psi|K(j)|\psi\rangle = |a|^2 v_n + |b|^2 v_{n'} + 2\mathrm{Re}(a^* b\, v_{nn'} e^{-i(k_n-k_{n'})ja}) \tag{5.79}$$

must not depend on j, and since we have assumed $k_n \neq k_{n'}$ when $n \neq n'$, off-diagonals in (5.78) will vanish. If some subbands $E_n(k)$ are degenerate, it is possible to find a basis of modes $\{\phi_m\}_m$ which diagonalizes (5.78).

Next, we specialize to propagating modes of the same direction, $\phi_{n,<}$ and $\phi_{m,<}$. We use the fact that the modes $\phi_{m,<}$ are eigenstates of the transfer matrix $g^R H_{-1}$ with eigenvalue $e^{ik_{m,<}a}$, where g^R is the retarded Green's function of the semi-infinite lead, see Eq. (5.89) of the next section, and we further use the definition of the lead's Γ matrix in terms of the self-energies, $\Gamma = i(H_1 g^R H_{-1} - H_1 g^A H_{-1})$, to find

$$\langle\phi_{n,<}|K(j)|\phi_{m,<}\rangle = \phi_{n,<}^\dagger \frac{ia}{\hbar}\left(H_1 g^R H_{-1} - H_1 g^A H_{-1}\right)\phi_{m,<} e^{i(k_{m,<}-k_{n,<})ja} \tag{5.80}$$

$$= \frac{a}{\hbar}\phi_{n,<}^\dagger \Gamma \phi_{m,<} e^{i(k_{m,<}-k_{n,<})ja}. \tag{5.81}$$

Therefore, modes are orthogonal with respect to the lead-current operator Γ,

$$\frac{a}{\hbar}\phi_{n,<}^\dagger \Gamma \phi_{m,<} = \delta_{n,m} v_{n,<}. \tag{5.82}$$

An identical relation holds for left-propagating modes $\phi_{n,>}$ and $\phi_{m,>}$. If we set $v_n = 0$ for evanescent modes, the relation holds for those as well, i.e. the operator Γ projects only onto propagating modes. In contrast to the orthogonality relation (5.79) with $K(j)$, we need modes of the same propagation direction in (5.82), because a representation of g^R was used which depends on the direction of propagation. In the case of degenerate subbands where $k_n = k_{n'}$ is possible for $n \neq n'$, a mode basis that diagonalizes Γ should be chosen.

We will need the generalized orthogonality relation (5.82), in order to prove the Fisher-Lee relation (5.74), by projecting onto modes in the asymptotic scattering states. We consider the eigenstate $\psi_{n,>}^{(q)}$ which originates from lead q, mode n, and we only care about propagating modes. We choose coordinates $x_p = 0$ and $x_q = 0$ of the first slices of leads p and q. The complex phase of the transmission and reflection coefficients $t_{pm,qn}$ will depend on the positions x_p and x_q, but this will not have any influence on physical results. At $x_p = 0$, the eigenstate can be written in modes of lead p, like $\psi_{n,>}^{(q)}(x_p) = \sum_m t_{pm,qn}|v_m|^{-\frac{1}{2}}\phi_{m,<}^{(p)}$, so the transmission amplitudes can be obtained from the orthogonality relation in lead p. For symmetry reasons, the same transmission amplitude is obtained if one considers a state $\psi_{m,<}^{(p)}$ which leaves only by mode m of lead p, and comes in by any mode of any lead. In this case, the orthogonality relation in lead q is needed. Together, we have

$$t_{pm,qn} = |v_{m,<}|^{-\frac{1}{2}}\phi_{m,<}^{(p)\dagger}\Gamma_p \psi_{n,>}^{(q)}(x_p) = \psi_{m,<}^{(p)\dagger}(x_q)\Gamma_q \phi_{n,>}^{(q)}|v_{n,>}|^{-\frac{1}{2}}. \tag{5.83}$$

Our discussion was for the transmission amplitudes. For the reflection amplitudes $t_{pm,pn}$, an additional term $\phi^{(p)\dagger}_{m,<}\Gamma_p\phi^{(p)}_{n,>}$ must be included, since the orthogonality relation does not make any statement about the overlap of modes with opposite directions of propagation. We still need expressions for the scattering states. They are given by the propagator equations

$$\psi^{(q)}_{n,>}(x_p) = iG^R(x_p,x_q)\Gamma_q\psi^{(q)}_{n,>}(x_q) \tag{5.84}$$

$$\psi^{(p)}_{m,<}(x_q) = \frac{1}{i}G^A(x_q,x_p)\Gamma_p\psi^{(p)}_{m,<}(x_p) \tag{5.85}$$

where G^R and G^A are the Green's functions of the combined device (sample and leads).

To prove the advanced propagator equation (5.85), it is sufficient to show that $G^A(x_q,x_p)$ fulfils the Schrödinger equation in x_q (which is clear by (5.9)), and that the correct initial condition is fulfilled. Assuming x_p to be in the asymptotic region of lead p, where $\psi^{(p)}_{m,<}(x_p) = \phi^{(p)}_{m,<}\frac{1}{\sqrt{|v_{m,<}|}}e^{ik_{m,<}x_p}$ is given by a single propagating (outgoing) mode $\phi_{m,<}$, the initial condition should be

$$\phi_{m,<} = \frac{1}{i}G^A(x_p,x_p)\Gamma_p\phi_{m,<}. \tag{5.86}$$

Proving this requires a bit more work. We use the Green's function of slice H_0 at coordinate x_p. Since x_p is in the asymptotic region of lead p, it is sufficient to take the Green's function of an infinite lead, consisting of left and right semi-infinite leads attached at the slice described by H_0. This is given by $G^A(x_p,x_p) = (E - H_0 - \Sigma^A_L - \Sigma^A_R)^{-1}$, where $\Sigma^A_{L/R}$ are the left/right semi-infinite lead's advanced self-energies. Instead of (5.86), we show the equivalent equation

$$i\phi^\dagger_{n,<}(E - H_0 - \Sigma^A_R - \Sigma^A_L)\phi_{m,<} = \phi^\dagger_{n,<}\Gamma_p\phi_{m,<} \tag{5.87}$$

which holds for any two propagating modes $\phi_{n,<}$ and $\phi_{m,<}$, but not for decaying modes. To evaluate the left side of (5.87), we use the quadratic eigenvalue problem (5.66) in the form $(E - H_0)\phi_{m,<} = (H_1\lambda_{m,<} + H_{-1}\lambda^{-1}_{m,<})\phi_{m,<}$. Further, we use the transfer properties (5.89) and (5.96), and the definition of self-energies $\Sigma^R_R = H_1 g_R H_{-1}$ and $\Sigma^A_L = H_{-1}g^A_L H_1$, to get $(E - H_0)\phi_{m,<} = (\Sigma^R_R + \Sigma^A_L)\phi_{m,<}$, which proves (5.87), having in mind the definition $\Gamma_p = i(\Sigma^R_R - \Sigma^A_L)$.

5.7 Lead's Green's function

We have already made use of the semi-infinite lead's Green's function g^R_p, which was needed to express the lead's self-energy $\Sigma^R_p = \tau_p g^R_p \tau^\dagger_p$, but we have not yet explained how to calculate it. In this section we will discuss an efficient and general method [Wim08] to obtain g^R_p and Σ^R_p, which also proves to be numerically stable. g^R_p is difficult to obtain directly. It is much easier to first consider a lead of infinite extension, and then construct g^R_p from the modes ϕ_n of the infinite system. This can be done by using a Dyson equation that relates the Green's functions of the semi-infinite system to that of the infinite system, or it can be done in a constructive approach, directly in terms of the modes of the infinite lead, supplemented with boundary conditions corresponding to the semi-infinite system [San+99; Run+08]. We take the latter approach. For Σ^R_p we only need the surface part of g^R_p, i.e. the submatrix with couples to the sample via τ_p. In our convention the semi-infinite lead extends from coordinates $x = 0$ to infinity. Here we use

the symbol $g_R = (g_p^R)_{0,0}$ for the surface part of a lead attached at the right edge of the sample, dropping the lead index p.

When we introduced retarded and advanced Green's functions, we included an infinitesimal energy shift $\eta = 0^+$, as in $g_\infty^R = (E + i\eta - H_l)^{-1}$, and we did not include this shift in the eigenproblem (5.66). Considering the inverse function of $E_n(k)$, $k_n(E)$, we recognize that the infinitesimal energy shift translates to a momentum shift in the way

$$k_n' = k_n(E + i\eta) = k_n(E) + i\eta \frac{\partial k_n(E)}{\partial E} = k_n + i\eta \frac{1}{\hbar v_n}. \tag{5.88}$$

Since for $x \neq x'$, $(g_\infty^R)_{x,x'}$ obeys the Schrödinger equation, the shift in k_n will tell which direction of propagation is still normalizable and may be used for the construction of g_∞^R in the region $x > x'$ or $x < x'$. A continuity condition can be used to find the diagonals of g_∞^R. Note, that sign(η) is only important for selection of the direction of propagation, and if we do this by different means, we do not need to include the parameter η in the numerical calculation of modes. Since eigenvalues are analytic functions of the matrix coefficients[2], the limit $\lim_{\eta \to 0\pm}$ is irrelevant for the set of eigenvalues $\{\lambda_n\}_n$ found as solutions of (5.66). The construction of the surface part of the Green's function of a semi-infinite lead for the right side of the sample can be written in terms of a transfer matrix that involves the solutions ϕ_n. All of the N right-propagating or right-decaying (i.e. outgoing) modes $\phi_{n,<}$ are needed, as can be seen by a formulation in transfer matrices [Run+08]. The transfer matrix for propagation to the right is given by the $N \times N$ matrix

$$g_R H_{-1} = U_< \Lambda_< U_<^{-1} \tag{5.89}$$

where the matrix $U_< = (\phi_{1,<}, ..., \phi_{n,<})$ consists of right-going modes and $\Lambda_<$ is the corresponding diagonal matrix of eigenvalues. For a lead p attached to the right of the sample, we have $H_1 = \tau_p$, so its self-energy is

$$\Sigma_R^R = H_1 U_< \Lambda_< U_<^{-1}. \tag{5.90}$$

Similar relations can be obtained for the left semi-infinite lead's Green's function, which involve left-decaying and left-propagating modes. The transfer matrix is $H_{-1} g_L = U_> \Lambda_> U_>^{-1}$, and the coupling to the sample is $\tau_p = H_{-1} = H_1^\dagger$, so

$$\Sigma_L^R = U_> \Lambda_> U_>^{-1} H_1. \tag{5.91}$$

In order to understand better the construction of the transfer matrix, suppose that we know that the transfer matrix has an eigenvector u,

$$g_R H_{-1} u = \lambda u. \tag{5.92}$$

Then, λ and u are also eigenvalue and eigenvector of the quadratic eigenvalue problem (5.66) for the given energy E. To show this, we first note that due to translational invariance, the right semi-infinite surface Green's function is invariant if an extra slice is attached at the beginning (compare Eq. (5.30), replacing $H_S \to H_0$ for the extra lead slice),

$$g_R = (E - H_0 - H_1 g_R H_{-1})^{-1}. \tag{5.93}$$

[2] More precisely, this statement holds only as long as no points of degeneracy are reached.

Using (5.92) and (5.93), we find

$$H_{-1}u = g_R^{-1}\lambda u = (E - H_0 - H_1\lambda)\lambda u \tag{5.94}$$

which proves that we must have $u = \phi_n$ and $\lambda = e^{ik_n a}$ for some index n. In a similar manner, we find transfer matrix expressions for a left semi-infinite lead, involving left-propagating and left-decaying modes,

$$H_{-1}g_L^R = U_> \Lambda_> U_>^{-1}, \quad g_L^R H_1 = U_> \Lambda_>^{-1} U_>^{-1}. \tag{5.95}$$

While the selection of decaying modes is fixed by geometry and is the same for retarded and advanced Green's functions, the latter involve propagating modes of the opposite direction. Therefore, we find that only for propagating modes $\phi_{m,<}$,

$$g_L^A H_1 \phi_{m,<} = \lambda_{m,<}^{-1}\phi_{m,<}. \tag{5.96}$$

Differently from the discussion given above, in our numerical implementation the left, top, and bottom lead's Green's functions are obtained by applying a rotation to the right lead's Green's function. This is an advantage in performance, because often the setup is symmetric and the same lead's Green's function can be used in different places. An even more important advantage appears in the treatment of (perpendicular) magnetic fields, which are included in the Hamiltonian by a vector potential. A unique choice of the vector potential is required for all parts of the sample, including leads. The discretization prescription (2.35) is used to map a continuum model onto a lattice model. However, the lead Hamiltonian (5.63) was required to be translationally invariant along the direction of the lead's extension, and this requirement cannot be met for leads with orthogonal orientation. E.g. the gauge $\mathbf{A} = (-By, 0, 0)$ allows for translationally invariant Hamiltonians of left and right leads, but for the top and bottom leads, the gauge needs to be changed to $\mathbf{A} = (0, Bx, 0)$. Our method is to supplement the rotation operation by a gauge transformation. Since the gauge transformation $\mathbf{A} \to \mathbf{A} + \nabla f(\mathbf{r})$ for the vector potential is equivalent to a basis transformation of the lattice site basis, it may be implemented by the operation

$$(g_p)_{ij} \to (g_p)_{ij} e^{i\frac{q}{\hbar}(f(\mathbf{r}_i) - f(\mathbf{r}_j))}. \tag{5.97}$$

If the Hamiltonian includes spin, the lead rotation procedure also includes a rotation on the spin space, $(g_p)_{ij} \to e^{-i\alpha J_z}(g_p)_{ij}e^{i\alpha J_z}$ for a lead that is rotated by angle α, where J_z is the generator for rotations on the spin space. The correctness of the numerical transport implementation is verified by checking symmetries of the effective Hamiltonian $H + \sum_p \Sigma_p$ that includes the leads, and by shifting the interface between sample and leads, which should not change any numerical output.

Formally, (5.89) provides a solution for calculating g_R and the lead's self-energy. However in practice, we may experience some numerical problems. The easiest way to proceed, is to first numerically solve the eigensystem (5.67). The numerical output will be a set of normalized vectors $(\phi_n, \lambda_n\phi_n)$. Since we take only the upper half components, we need to renormalize ϕ_n. Next, we calculate the velocities v_n according to (5.69), and use the λ_n and the v_n to sort out the N right-going solutions, which make up the columns of $U_<$. The next step would be the numerical matrix inversion for $U_<^{-1}$, but this can fail dramatically in certain cases. As mentioned earlier, the $\phi_{n,<}$ constitute an orthonormal set only if the $\phi_{n,k}$ do not depend on k. This is the case, if the lead Hamiltonian can be written in the way $H_{l\infty}(k) = E(k) + H_y$, where H_y does

not depend on k. The simplest example of this kind is $H_{l\infty}(k) = \frac{\hbar^2}{2m}(k^2 + \hat{k}_y^2) + V(y)$ with some transverse confinement potential $V(y)$. In contrast, in a strong perpendicular magnetic field, the propagating modes will be localized around a transverse coordinate $y_n = y_n(k)$. One observes [Rot09; Wim08] that with increasing magnetic field, the transverse modes $\phi_{n,<}$ become more and more linear dependent. The linear dependency can be quantified by the condition number $\mathrm{cond}(U_<)$.

The condition number of a matrix A is defined as $\mathrm{cond}(A) = ||A||\,||A^{-1}||$, where $||A|| = \max_{\mathbf{x}\neq 0}\frac{|A\cdot\mathbf{x}|}{|\mathbf{x}|}$ is the common matrix norm. $\mathrm{cond}(A)$ gives a measure of the expected factor by which the error of solution vector $\mathbf{x}$ will increase, with respect to the error of the right side in $A\cdot\mathbf{x} = \mathbf{b}$. The numerical error in the entries of $U_<^{-1}$ can be estimated as $\approx 10^{-15}\mathrm{cond}(U_<)$ due to the limited machine precision of the common `double` data type. $\mathrm{cond}(A)$ can be calculated from the singular value decomposition (SVD), $A = U\Sigma V^\dagger$, where U and V are unitary matrices and $\Sigma = \mathrm{diag}(\sigma_1, ..., \sigma_N)$ is a matrix with the so-called singular values $\sigma_i \geq 0$ on the diagonal. The SVD exists for any matrix (even if non-quadratic), and can be seen as generalization of a diagonalization. We have $\mathrm{cond}(U) = \mathrm{cond}(V) = 1$ since U, V are unitary, and it is easy to see that $\mathrm{cond}(A) = \frac{\max_i(\sigma_i)}{\min_i(\sigma_i)}$. If the columns of $U_<$ become more and more linear dependent, the condition number diverges.

Therefore, we have implemented an alternative way to calculate g_R, as proposed in [Wim08]. The basic idea relies on the observation that the expression $U_<\Lambda_<U_<^{-1}$ only depends on the invariant subspace spanned by the N vectors $(\phi_{n,<}, \lambda_{n,<}\phi_{n,<})$ of $2N$ components. Therefore, it is not necessary to calculate the precise eigenvectors of the $2N \times 2N$ matrix W in (5.67). It is sufficient to know the subspace that they span. Of course, if we are interested in the full S-matrix given by (5.74), we need all individual eigenvectors. But even then, it turns out that calculating the Green's functions with the method of the invariant subspace drastically reduces numerical errors. Also, for the S-matrix only propagating modes are needed, while also evanescent modes are needed for construction of g_R.

The most general basis transformation mixing only the '$<$' subspace can be written in terms of an invertible $N \times N$ matrix X,

$$U_1' = U_1 X \quad \text{where} \quad U_1 = \begin{pmatrix} U_< \\ U_<\Lambda_< \end{pmatrix} = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix}. \tag{5.98}$$

The invariance of formula (5.89) can be seen by

$$(g_R H_{-1})' = U_{21}'(U_{11}^{-1})' = U_{21}XX^{-1}U_{11}^{-1} = U_{21}U_{11}^{-1} = U_<\Lambda_<U_<^{-1} = g_R H_{-1}. \tag{5.99}$$

The goal is to find a good basis transformation X that eliminates the conditioning problem of $U_<$. Unitary or orthogonal matrices are optimal in the conditioning sense. A representation of W in terms of a unitary matrix Q and an upper triangular matrix T can be found by the Schur decomposition, which is given by

$$W = QTQ^\dagger. \tag{5.100}$$

The Schur decomposition works for any quadratic matrix, and is readily available in the LAPACK linear algebra package [And+99]. Further, one has the freedom of choosing any order of the diagonals of T, i.e. the eigenvalues λ_n, as long as these are non-degenerate. (The case of troublesome degeneracy only appears if one accidentally hits exactly a band crossing, which is

very unlikely in a numerical parameter sweep.) We write T in the form of $N \times N$ blocks, and find X_1 which diagonalizes the block T_{11},

$$T = \begin{pmatrix} T_{11} & T_{12} \\ & T_{22} \end{pmatrix} = \begin{pmatrix} X_1 & \\ & \mathbb{1} \end{pmatrix} \begin{pmatrix} D_1 & X_1^{-1}T_{12} \\ 0 & T_{22} \end{pmatrix} \begin{pmatrix} X_1^{-1} & \\ & \mathbb{1} \end{pmatrix}. \tag{5.101}$$

Combining the expressions for W and T, we find

$$WQ \begin{pmatrix} X_1 \\ 0 \end{pmatrix} = Q \begin{pmatrix} X_1 D_1 \\ 0 \end{pmatrix}. \tag{5.102}$$

We assume that the Schur decomposition is chosen so that the set of $\lambda_{n,<}$ is found on the diagonals of T_{11} and D_1. Then (5.102) shows that the corresponding eigenvectors of W are given by the columns of $Q_1 X_1$, where $Q = (Q_1, Q_2) = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$. As shown above in (5.99), the basis transformation X_1 does not modify the result, $g_R H_{-1} = Q_{21} Q_{11}^{-1}$. Here Q is unitary, so $\mathrm{cond}(Q) = 1$. Although this does not imply that $\mathrm{cond}(Q_{11})$ is also unity, in the examples that we treated numerically we find that $\mathrm{cond}(Q_{11})$ is of the order of unity, thus solving the numerical problem associated with the matrix inversion.

Note that throughout this section, we have assumed that $U_<$ is invertible. We are not aware of a general proof of this statement. However, in the examples that we have treated numerically, this seems to be the case, even though $U_<$ may become quite ill-conditioned in the case of a perpendicular magnetic field.

Chapter 6

Short introduction to topological insulators

The classification of different phases of matter, and finding principles that allow to distinguish between them sharply, has always been a fundamental question of interest for physicists.

In the context of the theory of phase transitions after Landau [Lan37], the classification is done by finding the difference in the fundamental symmetries of the system, an example being the discrete translational symmetry in a crystal, which is not present in the liquid or gaseous phases. A parameter that becomes non-zero when a certain symmetry is broken, is called order parameter. Consider for example a ferromagnet. When we lower the temperature below the Curie temperature, the rotational symmetry is spontaneously broken by the direction of the non-zero magnetization. Thus, the magnetization is an order parameter. The order parameter can be used to parametrize the thermodynamic free energy. If the latter has more than one local minimum, changing the order parameter can switch the global minimum of the free energy and thereby control the phase change.

But symmetry does not provide the only principle by which one can distinguish different phases of matter. There are also phases for which the distinguishing property lies in the topology of an associated mathematical function, e.g. in the case of a crystal the topology of the Bloch Hamiltonian $H(\mathbf{k})$, which is a function on the (magnetic) Brillouin zone. Historically, the first example of a topological classification of states is the quantum Hall effect (QHE), which was discovered experimentally in 1980 [Kli+80] in a two-dimensional electron gas (2DEG) with high electron mobility. A mathematical model explaining the precise quantization of the Hall conductance[1] by means of topology has been given in 1982 [Tho+82]. In a later work, Kohmoto [Koh85] gave a more accessible description of the relation of the QHE and topology.

A generic feature, that is required for classification of topologically non-trivial phases, is a gap in the bulk energy spectrum.[2] The reason is that topological invariants can be written in terms of integrals over (generalized) Berry curvatures. However, a non-trivial Berry phase can only appear if there is a projection on a subset of the Hilbert space (see Chapter 4 about the adiabatic theorem). Physically, in a gapped insulating system, the projection is given by the fact that the valence band is completely occupied, and usually transport is possible only by

[1] Usually, one should distinguish between the microscopic property conductivity and the macroscopic property conductance, but in a 2D system the units are the same.

[2] There also are gapless systems that are called topological, like e.g. Weyl semi-metals, but we will not be concerned with those.

excitation of electrons to the conduction band.[3] In the case of the QHE, the energy spectrum in the magnetic field is given by the discrete set of Landau levels, so this system also has a gapped spectrum. Thus, one will be interested only in insulators, when looking for time reversal invariant systems with non-trivial topology. Another generic feature of topological phases are edge states, which have energies usually lying in the bulk gap of the insulator. They always appear, when a topologically non-trivial system faces a trivial system (or vacuum) - this is called the bulk-boundary correspondence principle. The edge states are also responsible for the interesting features seen in the electronic transport. Having said so much about topologically interesting systems, we should also define what a topologically trivial system is, in order that we can distinguish the phases. A system is topologically trivial, if it is adiabatically connected to the atomic limit, without closing the bulk gap at any point during the parameter change. One may imagine the bonds between the atoms to elongate until the interaction is negligible, while tracking the evolution of the energy spectrum.

6.1 Quantization of the Hall conductivity

Let us outline the main ideas as discussed by Kohmoto [Koh85] for the QHE, since this will also help to understand the quantum spin Hall effect (QSHE) [Kan+05a; Ber+06a], which is a more recently discovered time reversal invariant topological phase that we will discuss later. For the quantum Hall effect, we consider the (single particle approximation) Hamiltonian of an infinite 2DEG on a lattice in presence of a perpendicular magnetic field $\mathbf{B}$,

$$H = \frac{(\mathbf{p} + e\mathbf{A})^2}{2m} + U(\mathbf{r}), \tag{6.1}$$

where $\mathbf{p} = -i\hbar(\partial_x, \partial_y, 0)$ and $U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R})$ is the lattice-periodic potential, $\mathbf{R}$ being some lattice vector. Although the system is translationally invariant, the vector potential $\mathbf{A}$ breaks the translational symmetry of the Hamiltonian. In spite of this, one can still make use of the translational symmetry and the Bloch theorem of quantum mechanics, by introducing the magnetic translation operators [Zak64]

$$\hat{T}_{\mathbf{R}} = e^{\frac{i}{\hbar}\mathbf{R}\cdot(\mathbf{p}-e\mathbf{A})}, \tag{6.2}$$

with the symmetric gauge $\mathbf{A} = \frac{1}{2}\mathbf{B}\times\mathbf{r}$. Then, $[p_i + eA_i, p_j - eA_j] = 0$. The magnetic translation operators act just as the usual translation operators on spatial functions, $\hat{T}_{\mathbf{R}}U(\mathbf{r}) = U(\mathbf{r}+\mathbf{R})\hat{T}_{\mathbf{R}}$, and one finds

$$[H, \hat{T}_{\mathbf{R}}] = 0 \tag{6.3}$$

even for non-zero magnetic field. Further, if we restricts ourselves to magnetic fields that give a rational number p/q of flux quanta per unit cell area ab, we can enlarge the unit cell by the factor q to define a magnetic unit cell enclosing an integer flux number. The magnetic translation operators $\hat{T}_{qa}$ and $\hat{T}_b$, which correspond to the magnetic unit cell's extension, commute and can be used to find simultaneous eigenfunctions of $H, \hat{T}_{qa}$ and $\hat{T}_b$. These are the generalized Bloch states $\psi^{\alpha}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{kr}}\langle\mathbf{r}|u^{\alpha}_{\mathbf{k}}\rangle$, where α is a band index. They are parametrized by momenta $\mathbf{k} = (k_x, k_y) \in [-\frac{\pi}{qa}, \frac{\pi}{qa}] \times [-\frac{\pi}{b}, \frac{\pi}{b}]$ in the magnetic Brillouin zone (MBZ), which is a torus

[3] The principle of transport that generates the transverse current in the QHE is different, because it does not rely on charge excitations. Instead, the transport can be understood in terms of an adiabatic shift of the transverse coordinate of bulk states, which can be shown to be localized in transverse direction, if an appropriate gauge for vector potential is assumed [Lau81; Hal82].

$T^2 = S^1 \times S^1$ because of the periodic boundary conditions in $\mathbf{k}$. The complex phase of the wave functions $|u^\alpha_\mathbf{k}\rangle$, seen as function over the MBZ, defines a fibre bundle, where the fibre is the $U(1)$ group of the complex phase, which is attached to each point of the MBZ. The key insight is, that the quantized value of the Hall conductance depends only on the topology of the fibre bundle, i.e. the way that it is internally twisted. We will see that if there is a zero $\langle \mathbf{r}_0 | u^\alpha_{\mathbf{k}_0}\rangle = u^\alpha_{\mathbf{k}_0}(\mathbf{r}_0) = 0$, the fibre bundle can have a non-trivial topology.[4] The existence of such a zero is guaranteed by the non-zero flux per magnetic unit cell, as can be shown by integrating the complex phase change of $u_{\mathbf{k}_0}(\mathbf{r})$ over the closed loop given by the boundary of the magnetic unit cell.

The Hall conductivity σ_{xy} at zero temperature, in linear response to an applied electric field, is given by Kubo's formula [Kub57]. We also give a derivation by linear response in Appendix A, since the equivalence of the formula originally given by Kubo and the form in which it is cited by Kohmoto,

$$\sigma_{xy} = \frac{e^2\hbar}{i} \sum_{E_n < E_F < E_m} \frac{(v_y)_{nm}(v_x)_{mn} - (v_x)_{nm}(v_y)_{mn}}{(E_n - E_m)^2}, \tag{6.4}$$

is not trivial. Here, indices n, m identify the eigenstates of energy E_n, E_m and should be understood as collective indices including the band α and momentum $\mathbf{k}$. The sum is over all pairs of occupied states n and unoccupied states m, so it captures the physics of virtual excitations. Matrix elements of the velocity operator $\hat{\mathbf{v}} = \frac{\partial H}{\partial \mathbf{p}}$, which is proportional to the paramagnetic current, are denoted by $(v_{x/y})_{nm}$. In the momentum basis, the derivative $\nabla_\mathbf{k}$ in $\hat{\mathbf{v}}$ can be shifted to the states $|u^\alpha_\mathbf{k}\rangle$, and the contribution to the Hall conductivity of a completely filled band α can be written as an integral over the Berry curvature,

$$\sigma^\alpha_{xy} = \frac{e^2}{h}\frac{1}{2\pi}\int_{\mathrm{MBZ}} dk^2 \left(\nabla_\mathbf{k} \times \hat{\mathbf{A}}^\alpha(\mathbf{k})\right)_3 \tag{6.5}$$

with the Berry connection associated to the $U(1)$ bundle over the MBZ given by

$$\hat{\mathbf{A}}^\alpha(\mathbf{k}) = \int_0^{qa} dx \int_0^b dy\, u^{\alpha *}_\mathbf{k}(x,y)\frac{1}{i}\nabla_\mathbf{k} u^\alpha_\mathbf{k}(x,y) = \frac{1}{i}\langle u^\alpha_\mathbf{k}|\nabla_\mathbf{k}|u^\alpha_\mathbf{k}\rangle, \tag{6.6}$$

and $\nabla_\mathbf{k} \times \hat{\mathbf{A}}^\alpha(\mathbf{k})$ being the Berry curvature. The integral (6.5) is also known as the first Chern number (times $\frac{e^2}{h}$) of the fibre bundle, and it is known to be a topological invariant, meaning that it captures features which are invariant under smooth deformations of the fibre bundle [Nak03]. It depends only on the kind the fibre is internally twisted and whether it has "holes". At this point, the attentive reader might ask why the Chern number should be a function of only the $U(1)$ bundle - seemingly it depends on the whole function $u^\alpha_\mathbf{k}(x,y)$ and not just its phase. The following steps will show that only the phase matters. It is easy to see that the Berry curvature is invariant under a gauge transformation $|u^\alpha_\mathbf{k}\rangle' = e^{if(\mathbf{k})}|u^\alpha_\mathbf{k}\rangle$ with some arbitrary but smooth function $f(\mathbf{k})$. In this sense, it is analogue to the electromagnetic field tensor, which is invariant under gauge transformations in real (as opposed to reciprocal or momentum) space.

By Stokes' theorem, (6.5) must evaluate to zero if $\hat{\mathbf{A}}^\alpha(\mathbf{k})$ smooth function with a global single-valued definition, since the MBZ is a torus and does not have a boundary. The Berry

[4] Although such a zero is required for a non-trivial topology and thus for a non-zero value of σ_{xy}, from Kohmoto's argument we do not find a simple relation between the number of roots $\mathbf{k}_i$ with $u_{\mathbf{k}_i}(\mathbf{r}_0) = 0$ in the MBZ, and the integer value $\sigma_{xy}\frac{h}{e^2}$.

connection transforms under the gauge change like $\hat{\mathbf{A}}'^{\alpha}(\mathbf{k}) = \hat{\mathbf{A}}^{\alpha}(\mathbf{k}) + \nabla_{\mathbf{k}} f(\mathbf{k})$. So the complex phase of the states is essential for the value of the Berry connection. A convention to define the phase of $|u^{\alpha}_{\mathbf{k}}\rangle$ for all $\mathbf{k} \in$ MBZ is needed. To this end, one can choose an arbitrary $\mathbf{x}_0$ in the magnetic unit cell and adopt a phase convention which defines that $\langle \mathbf{x}_0 | u^{\alpha}_{\mathbf{k}} \rangle$ should be real for all $\mathbf{k}$. But there is a difficulty - as already mentioned above, for non-zero magnetic field, there will always be some $\mathbf{k}_0$ where $\langle \mathbf{x}_0 | u^{\alpha}_{\mathbf{k}_0} \rangle = 0$, so there, the phase of $|u^{\alpha}_{\mathbf{k}_0}\rangle$ must be defined in a different way. This is possible by choosing a different reference point $\mathbf{x}_1$ instead of $\mathbf{x}_0$, for the phase convention in the vicinity of $\mathbf{k}_0$. Like this, one can evaluate the Chern integral (6.5) using two (or more) overlapping patches, on which the Berry connection is defined within different gauges. The situation is analogue to the hypothetical magnetic monopole discussed by Dirac [Dir31]. The value of the integral does not depend on the way the MBZ is divided into patches, nor does it depend on the gauges used locally for the patches. However, it does depend on the transition function defined in the overlapping area of the patches. We will call the different patches that make use of $\mathbf{x}_0$ and $\mathbf{x}_1$, P_0 and P_1. In the overlap $P_0 \cap P_1$, Kohmoto defines the transition function $\chi(\mathbf{k})$ by

$$|u^{(1)}_{\mathbf{k}}\rangle = e^{i\chi(\mathbf{k})} |u^{(0)}_{\mathbf{k}}\rangle. \tag{6.7}$$

The transition function gives a local gauge change, as opposed to a global gauge change, which would be irrelevant for the integral. By Stokes' theorem, the Chern integral reduces to a line integral over the transition function,

$$n = \frac{1}{2\pi} \oint_{\partial P_1} d\mathbf{k} \cdot \nabla_{\mathbf{k}} \chi(\mathbf{k}). \tag{6.8}$$

The result must be an integer n, because the function $\chi(\mathbf{k})$ can only change by $2\pi n$ when we go around the loop of the line integral once, otherwise the complex phase of $|u^{\alpha}_{\mathbf{k}}\rangle$ couldn't be defined uniquely within each patch. The integer n is also often referred to as TKNN integer after [Tho+82]. Eq. (6.8) also proves that the Chern number depends only on the phase of the generalized Bloch functions, and allows us to identify the phase degree of freedom as $U(1)$ fibre bundle on a torus (the MBZ) as base manifold. In conclusion, it has been shown that the Hall conductivity σ_{xy} is given by a topological invariant, and is quantized to integer times $\frac{e^2}{h}$.

The derivation by Kohmoto, which we have outlined above in order to explain the exact quantization of the Hall conductivity, required a perfect crystal, and thus cannot explain the insensitivity to disorder. Disorder is important because in realistic physical systems, there are always imperfections in the lattice, like vacancies, dislocations, impurities, and other interactions and scattering mechanisms like phonons, none of which were included in the model discussed by Kohmoto. It is quite interesting that the experimentally observed quantum Hall resistance is insensitive to all these perturbations. The stability of the Hall signal must be related to the fact that it is a topological feature - as long as perturbations are not strong enough to close the gap of the bulk spectrum, the topological invariant will not change. Niu et al. [Niu+85] have shown how to reformulate the Chern number in terms of an integral over parameters controlling twisted boundary conditions. The main idea of the approach is that the topological invariant is a property of the bulk, so the effect of changing the boundary conditions of bulk states must be exponentially suppressed. Therefore, one can average over angles parametrizing the boundary condition. The averaging integral takes the place of the integral over the Brillouin zone in the Kubo formula. Disorder may be introduced, since one no longer needs to work with generalized Bloch states.

In all the arguments given above, a clear deficiency is that we did not yet consider the effect of edges, which are always present in a realistic sample, which must be finite. In a magnetic field perpendicular to the plane of the 2DEG, the Lorentz force will push the electrons, which contribute to transport, toward the edge. A quantum mechanical calculation [Hal82] shows that in a sufficiently strong magnetic field, edge states appear which are exponentially localized. Each edge state carries the current corresponding to a conductance quantum $\frac{e^2}{h}$. They are chiral, meaning that left- and right-propagating edge states are localized at opposite edges. This prevents backscattering even if disorder, e.g. impurities in the lattice are present, since it would require scattering between opposite sample edges, which are assumed to be far apart. In the vicinity of an impurity, an edge state will simply adapt its shape to go around the impurity. In this picture, the perfect quantization of the Hall conductance can be explained simply by counting the number of edge states. Regarding the fundamental difference between calculation of σ_{xy} for the bulk of an infinite system by the Kubo formula, and the argument based on counting the edge states, it seems astonishing that one obtains exactly the same Hall conductivity or conductance by both approaches. It is known that at the interface between two phases, which differ by the value of their topological invariant, the difference leads to the presence of this number of edge states. This is called the bulk-boundary correspondence principle [Vol09; Gur11].

So far, we have explained that the quantization of the Hall conductance may be understood either as property of the bulk or by counting the edge states. In order to understand that these explanations provide different viewpoints of the same physical behaviour, rather than separate effects that should be added up to give a combined Hall conductance, it is interesting to compare with an argument given by Laughlin [Lau81]. He does not calculate the Hall current by the Kubo formula, but rather generates a current by a periodic and adiabatic pumping procedure, which causes a transverse shift of the occupied states. A net effect in the occupation of electronic levels is only found at the edges of the sample, giving rise to the Hall effect. Although Laughlin discusses the non-equilibrium occupation that is obtained at the sample edges, he does not yet mention the exponentially localized edge states. In a slightly modified geometry with respect to Laughlin, Halperin [Hal82] discusses the importance of the edge states and also gives clearer arguments for the stability of the quantization in disordered samples.

In order to get a simple picture of the equivalence of the bulk Hall current and edge state Hall current, we consider what happens when we change the boundary conditions that apply for the current. Let us consider a 2DEG of rectangular shape as shown in Fig. 6.1a, with length L (x-direction) and width W (y-direction). A magnetic field $\mathbf{B} = B_z \mathbf{e}_z$ is applied perpendicularly to the plane of the 2DEG. Reservoirs (leads), labelled with capital letters, are attached to the left (A), top right (B) and bottom right (C). An electric field is applied in x-direction, $\mathbf{E} = E_x \mathbf{e}_x$. In Fig. 6.1a we identify the top and bottom edges of the sample by requiring periodic boundary conditions. This serves to model an infinite extension of the sample in transverse (y) direction. The artificial top and bottom edges of the sample are shown as dotted lines. The classical trajectory of an electron in this configuration is given by spiral with a constant transverse drift velocity,

$$\mathbf{r}(t) = \frac{1}{\omega_c} \left(R_z(\omega_c t - \frac{\pi}{2}) - R_z(-\frac{\pi}{2}) \right) \mathbf{v}_0 + \frac{E_x}{B_z} \mathbf{e}_y t + \mathbf{r}_0, \tag{6.9}$$

where R_z is the matrix for rotations about $\mathbf{e}_z$, and $\omega_c = \frac{eB}{m}$ is the cyclotron frequency. So within an infinite system or equivalently, a finite system with periodic boundary conditions, there is no transport in direction of the applied field at all. This still holds in a quantum mechanical system.

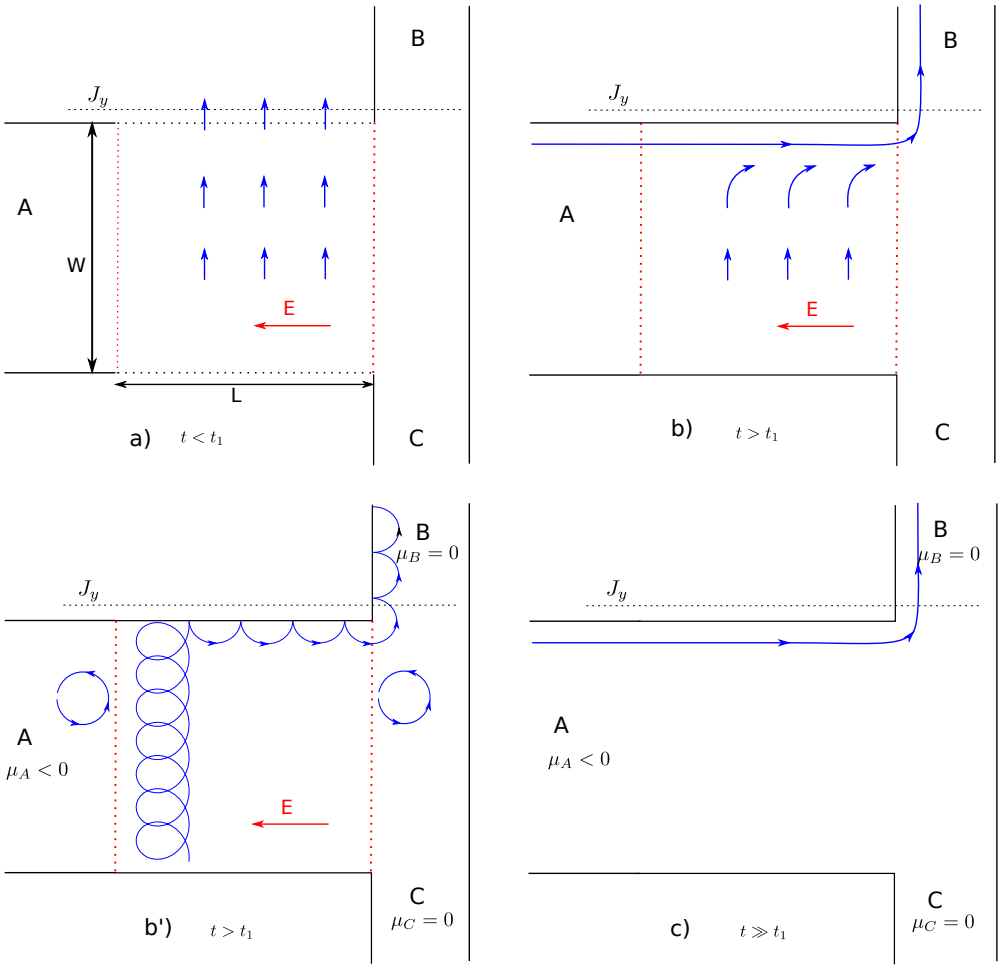


Figure 6.1: Illustration of the equivalence of bulk conductivity σ_{xy} and edge state conductance - compare discussion in the text. A 2DEG of length L and width W is subject to a perpendicular magnetic field and a homogeneous electric field $\mathbf{E} = -E\mathbf{e}_x$, which is non-zero in the region between the vertical dotted red lines. We define the Hall current as total transverse current J_y current passing the dashed line shown at the top (for clarity, the line is slightly shifted away from the edge of the sample). (a) For times $t < t_1$, we identify top and bottom edges in the region $0 \leq x \leq L$ via periodic boundary conditions, which effectively model an infinite transverse extension of the sample. In the stationary state, there is no longitudinal current. Note that this setup cannot be simulated within the Landauer-Büttiker formalism, because the latter does not include the electric field $\mathbf{E}$, but only potentials μ_l of the leads $l = A, B, C$. (b) At time t_1, the periodic boundary conditions are changed to hard-wall boundary conditions. Chiral edge states appear. The edge state at the top is shown in blue. The current distribution changes, since the transverse current is collected in the edge state. (b') For the purpose of better physical understanding, a sketch of the classical situation is shown, where the electron trajectories are spirals as given by (6.9), and reflections occur at the sample boundaries. Longitudinal transport exists only due to the boundaries. (c) At times $t \gg t_1$, the current distribution has become stationary. This setup can be modelled within the Landauer-Büttiker formalism, with the shown potentials μ_A and μ_B at the leads. The current J_y is carried only by the edge states, and its value is the same as in (a).

At some time t_1, we replace the periodic boundary conditions by hard wall boundary conditions, which are illustrated as solid lines at the sample edges in Fig. 6.1b. We define the Hall current as integrated current passing the dotted horizontal line (labelled with J_y) at the top edge of the sample. In the figures, this line is slighty shifted to the top for clarity. For times $t < t_1$ the Hall current is given by $J_y = \int_0^L dx\, j_y = \sigma_{yx} L E_x = \sigma_{yx} \Delta U_x$, where $\Delta U_x = (\mu_B - \mu_A)/e$ is the longitudinal voltage drop. If σ_{yx} is calculated from the Kubo formula (compare Appendix A), we obtain the quantized Hall conductance $\frac{J_y}{\Delta U_x}$. From the time t_1 on, when hard wall boundary conditions are introduced, the system shows the chiral edge states of the quantum Hall state. Beginning at t_1, the current J_y must completely collect into the edge states. We assume that the upper edge state carries current to the right. Lead B in Fig. 6.1b can be used to measure the Hall current J_y, which has been collected by the edge state. Some time later at $t \gg t_1$, a stationary state will be reached. Due to the boundary, electrons can no longer move up in the region $0 \leq x \leq L$. The vector field for the current distribution $\mathbf{j}(\mathbf{r})$ (sketched by blue arrows in Fig. 6.1c) looks completely different than for times $t \ll t_1$, pointing upward instead of toward the right. The bulk no longer contributes to the transverse current J_y, since the latter completely originates from the edge state. The quantized value of J_y has not changed while we have changed the system, showing the subtle equivalence between Hall conductance originating from the edge-states and Hall conductivity of an infinite 2DEG, as obtained by the Kubo formula. In short, we have exploited that the infinite system does not show longitudinal transport at all, and when hard wall boundaries are introduced, the current which is carried by the edge states must originate from redirecting the original transverse current at the boundaries. Since there is no backscattering, changing the system is not able to change the current J_y.

6.2 Two-dimensional topological insulators

The second example of a distinct new topological state of matter that has been discovered, is the two-dimensional (2D) topological insulator (TI), which does not require a magnetic field [Kan+05a; Kan+05b; Ber+06a; Kön+07; Kön+08; Rot+09]. Experimentally, the fingerprint of such a system is the quantum spin Hall effect (QSHE). Instead of a magnetic field, strong spin-orbit (SO) coupling is needed. In the case of S_z-conserving SO terms, this can formally be compared to a magnetic field acting on carriers of different spins separately, but with different signs [Ber+06b]. The appearance of the QSHE is due to a pair of helical edge states at each edge of the sample. Each edge state carries a conductance quantum. By helical, it is meant that the directions of momentum and spin are locked. This is different from the chiral edge states of the QHE. There, all the edge states which are localized at the same side of the sample, propagate in the same direction (the spin is not required in a model showing the QHE). Although time reversal symmetry is conserved, the 2D TI has topological features distinct from a trivial band insulator (which does not show the QSHE signal). Time reversal symmetry is very important for the theoretical classification of the state. By adding a perturbation that breaks time reversal symmetry, it is possible to open a gap inside the edge state spectrum, rendering the system equivalent to a trivial band insulator, and eventually destroying the QSHE signal. However, the gapless edge states are protected against small time-reversal symmetric perturbations. In this sense, it is said that time reversal symmetry is protecting the topological state, also showing that it is fundamentally different from the topological state of the QHE.

The protection is connected with Kramers' theorem, since the edge states are related by time reversal symmetry. More precisely, only an odd number of Kramers pairs is protected

against single-particle excitations [Kön+08]. To explicitly show the protection, let us denote the right-propagating edge state as $|\psi_+\rangle$. The left-propagating edge state at the same edge is related by time reversal, $|\psi_-\rangle = \mathcal{T}|\psi_+\rangle$. Using the property $\mathcal{T}^2 = -1$ of a spin-$\frac{1}{2}$ system, and the anti-unitarity of $\mathcal{T}$,

$$\langle \mathcal{T}\psi_1 | \mathcal{T}\psi_2 \rangle = \langle \psi_2 | \psi_1 \rangle, \tag{6.10}$$

we can show that a Hermitian perturbation V that conserves time reversal symmetry, $V^\dagger = V = \mathcal{T}V\mathcal{T}^{-1}$, does not allow for backscattering of the edge states,

$$\begin{aligned} \langle \psi_+ | V | \psi_- \rangle = \langle \psi_+ | V\mathcal{T} | \psi_+ \rangle = \langle \psi_+ | \mathcal{T}V | \psi_+ \rangle = \langle V\psi_+ | \mathcal{T}\psi_+ \rangle &\overset{(6.10)}{=} \langle \mathcal{T}^2\psi_+ | \mathcal{T}V\psi_+ \rangle \\ = -\langle \psi_+ | V | \mathcal{T}\psi_+ \rangle = -\langle \psi_+ | V | \psi_- \rangle &= 0. \end{aligned} \tag{6.11}$$

In an infinitely long wire, any finite backscattering would eventually suppress transport completely, opening a gap at the Fermi level. So time reversal conserving perturbations cannot open a gap in the edge state dispersion.

This principle also allows for a simple definition of the Z_2 topological invariant of a TI. Speaking of a Z_2 integer, it is meant that only even or odd matters. Let us split up the Hamiltonian of a system under consideration, into a part that conserves S_z, and a perturbation that couples opposite spins. One obtains topological invariants (Chern numbers) $n_\uparrow$ and $n_\downarrow$ for the decoupled systems of spin $\uparrow, \downarrow$, given by this number of edge states at a single edge, just as in the QHE case. Because of time reversal symmetry, the TKNN integer $n_\uparrow + n_\downarrow$ must vanish [Kan+05b], corresponding to the intuitive picture of opposite magnetic fields for spin $\uparrow$ and $\downarrow$. However, the Z_2 invariant, which is simply defined as $\frac{1}{2}(n_\uparrow - n_\downarrow) = n_\uparrow$, can be non-zero. Suppression of backscattering guarantees that the invariant remains unchanged, if the perturbation is included and is not too strong.

The QSHE has first been predicted theoretically in 2005, in a system of graphene with SO interactions that open gaps at the Dirac points [Kan+05a], and an explanation in terms of a Z_2 topological invariant has been given [Kan+05b]. The topological Z_2 invariant introduced in [Kan+05b] is more general than the definition in terms of Chern numbers, and does not require the perturbation connecting opposite spins to be small. However, its definition is more complicated and it is often not easy to evaluate. The general invariant may be calculated by counting the number of pairs of zeros in the BZ of the Pfaffian[5] $P(\mathbf{k}) = \mathrm{Pf}\left[\langle u_i(\mathbf{k}) | \mathcal{T} | u_j(\mathbf{k}) \rangle\right]$. Here, the antisymmetric matrix of overlaps of time-reversed Bloch states $\mathcal{T}|u_j(\mathbf{k})\rangle$ with other Bloch states is needed. Due to the smallness of the SO-induced gap in graphene ($\approx \mu eV$), graphene is not suitable for experimental detection of the phase. In 2006, the QSHE has been theoretically predicted to appear in a system of HgTe/CdTe quantum wells [Ber+06a]. Shortly thereafter, experimental detection of the QSHE [Kön+07] has proven the system to be a topological insulator. Near the interesting regime of the topological phase transition, the physics of this two-dimensional system is described by an effective 4×4 Hamiltonian for the lowest electron-like and heavy-hole levels of the quantum well (QW). The effective block diagonal four band Hamiltonian introduced by Bernevig, Hughes and Zhang (BHZ) [Ber+06a]

$$H(\mathbf{k}) = \begin{pmatrix} h(\mathbf{k}) & 0 \\ 0 & h^*(-\mathbf{k}) \end{pmatrix}, \quad h(\mathbf{k}) = \epsilon(k) + \begin{pmatrix} \mathcal{M}(k) & Ak_+ \\ Ak_- & -\mathcal{M}(k) \end{pmatrix} \tag{6.12}$$

[5] The Pfaffian of an antisymmetric matrix A is defined in a way that $\mathrm{Pf}(A) = \pm\sqrt{\det(A)}$, the extra information with respect to the determinant being the well-defined sign. If A is a 2×2 matrix, $\mathrm{Pf}(A) = A_{12}$.

consists of a modified 2D Dirac Hamiltonian $h(\mathbf{k})$ for one spin[6], let's call it spin $\uparrow$, and a block $h^*(-\mathbf{k})$ related to it by time reversal, for spin $\downarrow$. Here, $\mathcal{M}(k) = M - B(k_x^2 + k_y^2)$. The parameter B should not be confused with a magnetic field, which we do not include here. In the topologically non-trivial parameter regime where $M/B > 0$, this system shows helical pairs of edge states, which lead to the QSHE. It is interesting to note that quadratic terms ($\propto B$) are needed to define the topological regime. For $B = 0$, switching the sign of M simply amounts to a unitary operation. We will later discuss the material system of HgTe/CdTe quantum wells and derivation of the effective model in more detail.

In the block diagonal form of $H(\mathbf{k})$, the first Chern number is a topological invariant that can be separately evaluated for the blocks $h(\mathbf{k})$ and $h^*(-\mathbf{k})$. It can be obtained by integrating the Berry curvature over the Brillouin zone of a lattice model corresponding to (6.12). The lattice regularization $k_i \to \frac{1}{a}\sin(k_i a)$ is needed here, because the topological invariant is defined only for a compact base manifold, which is not contractible to a point. The lattice regularization cannot change the invariant, since the topological phase transition happens only in the vicinity of the Γ point, as can be seen by analyzing the limit $M \to 0$ [Bud+12b]. As already mentioned earlier, the Z_2 topological invariant of the topological insulator is then calculated as difference of Chern integers of the blocks.

At this point, we want to show in the simplest possible way, how to obtain the Chern number from a block of (6.12) for just one spin. Defining the vector $\mathbf{d} = (Ak_x, -Ak_y, \mathcal{M})$, the eigenenergies of (6.12) are conveniently expressed as

$$E_\pm(k) = \epsilon(k) \pm d(k) = \epsilon(k) \pm \sqrt{A^2k^2 + \mathcal{M}(k)^2}. \tag{6.13}$$

The corresponding normalized eigenvectors are given by

$$|u_+(\mathbf{k})\rangle = \frac{1}{\sqrt{2d(d+\mathcal{M})}} \begin{pmatrix} \mathcal{M}+d \\ Ak_-/k \end{pmatrix}, \quad |u_-(\mathbf{k})\rangle = \frac{1}{\sqrt{2d(d-\mathcal{M})}} \begin{pmatrix} \mathcal{M}-d \\ Ak_-/k \end{pmatrix}. \tag{6.14}$$

Let us consider the eigenvectors of positive energy $|u_+(\mathbf{k})\rangle$. We could also use the negative energy band, which corresponds to the occupied band in the semiconductor picture, but only the sign of the Berry curvature would change, and a sign change by definition does not change the Z_2 invariant. The Berry connection on the parameter space ($\mu = k_x, k_y, M$) is defined as $A_\mu^+ = -i\langle u_+(\mathbf{k})|\partial_\mu|u_+(\mathbf{k})\rangle$. Explicit calculation results in

$$(A_{k_x}^+, A_{k_y}^+) = \frac{A^2}{2d(d+\mathcal{M})}(k_y, -k_x). \tag{6.15}$$

We have already encountered this expression in a bit different form, when we discussed the semiclassical wave packet approach for the Dirac equation, in Section 4.2. Taking the Berry connection (4.41) and replacing $\mathbf{p} \to \mathbf{k}$, $p_z \to 0$, $c \to A$, $m_0c^2 \to \mathcal{M}$ and $E_p \to d$ reproduces our result (6.15). In the 2D problem, $\mathbf{A}^+$ becomes spin-diagonal and we need not worry about non-Abelian corrections. Next, we consider the Berry curvature, which is needed for the Chern integral. Direct evaluation gives

$$F_{xy} = \partial_{k_x} A_{k_y}^+ - \partial_{k_y} A_{k_x}^+ = \frac{1}{2}\hat{\mathbf{d}} \cdot (\partial_{k_x}\hat{\mathbf{d}} \times \partial_{k_y}\hat{\mathbf{d}}) \tag{6.16}$$

$$= -\frac{A^2(M + Bk^2)}{2d^3}. \tag{6.17}$$

[6] To be precise, $h(\mathbf{k})$ describes a coherent mixture of spins, with dominant contribution of spin up. This is a detail which will be discussed more thoroughly in Sections 7.1.1 and 10.6, and should not stop us from using labels $\uparrow, \downarrow$ here.

The expression (6.16) allows for a geometrical interpretation of the integral over the Berry curvature as the spatial angle that is subtended by the unit vector $\hat{\mathbf{d}} = \frac{\mathbf{d}}{d}$ on the sphere [Ber+06a], during integration over the base manifold of (k_x, k_y). If the base manifold is compact, the integral over the Berry curvature is a topological invariant, and the angle can change only by values of 4π. Explicitly, the invariance may be shown by considering the deformation by an arbitrary but smooth rotation matrix $R(\mathbf{k})$, which is applied to $\hat{\mathbf{d}}(\mathbf{k})$. Using $\det R(\mathbf{k}) = 1$, it is easy to check that $\hat{\mathbf{d}}'(\mathbf{k}) = R(\mathbf{k})\hat{\mathbf{d}}(\mathbf{k})$ gives the same Berry curvature (6.16). In our special case, the integration over a compact manifold may be replaced be an integration over the infinite 2D plane, without changing the value of the invariant. For $k = 0$, the vector $\hat{\mathbf{d}}$ points upward or downward depending on the sign of M, and for $k \to \infty$, $\hat{\mathbf{d}}$ also points upward or downward depending on the sign of B. Therefore, a lattice regularization is not required here. We obtain the Chern number for spin up,

$$\begin{aligned} n_\uparrow &= \frac{1}{2\pi}\int dk^2\, F_{xy} = \int_0^\infty dk\, k \frac{-\frac{M}{A}\cdot\frac{B}{A}k^2}{2\left(k^2 + (\frac{M}{A} - \frac{B}{A}k^2)^2\right)^{3/2}} \\ &= \frac{-\frac{B}{A}k^2 + \frac{M}{A}}{2\sqrt{k^2 + (\frac{M}{A} - k^2\frac{B}{A})^2}}\Bigg|_{k^2=0}^{\infty} = -\frac{1}{2}\left(\operatorname{sign}\frac{B}{A} + \operatorname{sign}\frac{M}{A}\right). \end{aligned} \tag{6.18}$$

Since the Chern number of the opposite spin evaluates to $n_\downarrow = -n_\uparrow$, the difference $(n_\uparrow - n_\downarrow)/2 = n_\uparrow$ is a topological invariant that is not cancelled by time reversal invariance. The Chern number (6.18) can take values $1, 0, -1$. However, exchanging spin up and spin down in the BHZ model simply corresponds to a unitary basis transformation. For this reason, the Z_2 invariant is only defined in terms of odd and even. If we identify values 1 and -1 of $n_\uparrow$ and use $\operatorname{sign}\frac{B}{M} = \operatorname{sign}\frac{B}{A}\operatorname{sign}\frac{M}{A}$, we obtain

$$n_\uparrow \bmod 2 = \frac{1}{2}\left(1 + \operatorname{sign}\frac{B}{M}\right) \tag{6.19}$$

independent of the sign of the parameter A.

In Chapter 7.1.1 we extend the derivation of Bernevig et al. [Ber+06a] to asymmetric QWs. Then, the structural inversion asymmetry leads to extra terms on the off-diagonals of (6.12). The simple procedure outlined above, obtaining the Z_2 invariant from two Chern integers, no longer works in this case. As mentioned above, the invariant is still well-defined in terms of the bulk Bloch states and the way they connect by the time reversal operator [Kan+05b]. Rather than analysing the bulk states, it is numerically much more straightforward to analyse the dispersion of a confined system. If the gapless helical edge states (or more precisely, an odd number of pairs of edge states at each edge) still exist in the bulk gap, the invariant is non-zero. In the block-diagonal model, each block is responsible for a single edge state per edge, and they are counter-propagating, since they are related by time reversal symmetry. A structure which lacks inversion symmetry, which could be due to structural (SIA) or bulk inversion asymmetry (BIA), allows for spin-orbit terms that conserve time reversal symmetry [Win05]. We have already seen that these terms do not allow for backscattering in the edge states of a single edge.

We can use a simple toy model to understand why edge states appear if one matches a trivial insulator with a topological (i.e. inverted-band) insulator. The vacuum can be considered as trivial insulator with the relativistic band gap $2m_0c^2$. Let us consider the simplest possible 2×2 Dirac Hamiltonian for just one spatial coordinate, $H_{D1} = \begin{pmatrix} M & Ak_x \\ Ak_x & -M \end{pmatrix}$. The spectrum is

$E_\pm = \pm\sqrt{A^2k_x^2 + M^2}$. Let us assume that the parameter M changes its sign at the interface $x = 0$, and let us look for possible solutions with an energy in the bulk gap $[-|M|, |M|]$. The eigenspinors are $\psi_\pm = (M \pm |E|, Ak_x)$. The two-component linear Dirac equation has only the boundary condition of continuity. Requiring spinors of the same energy E at left and right side of the interface to match, still allows for a sign change in k_x. One finds that both the sign of M and the sign of k_x has to change to form a continuous total wave function, and since this also corresponds to an eigenvector of $-H_{D1}$ with energy $-E$, a solution is possible only for $E = 0$. So $k_x = \pm i\frac{|M|}{|A|}$ is imaginary, and the evanescent solutions of the left and right side combine to form a cusp at $x = 0$. The fact that the edge states lie in the bulk gap forces them to be localized at the edge $x = 0$. By perturbation theory, we can also form an effective 1-band model quadratic in k. Then, the sign change of the gap parameter M corresponds to a sign change in the effective mass $m = M$. Since the model is quadratic, a second boundary condition involving the effective mass appears, $\frac{\partial_x\psi(x)}{M(x)}\Big|_{x=0^-} = \frac{\partial_x\psi(x)}{M(x)}\Big|_{x=0^+}$. This condition is responsible for the cusp feature in the effective model.

Introducing a second dimension in the toy model requires a bit more work. This would be needed to find the edge state dispersion which is linear in k_y. In particular, one also needs to include quadratic terms ($B \neq 0$) in the BHZ Hamiltonian (6.12) to have a well-defined topological regime. In [Kön+08] an explicit derivation for the linear edge state dispersion is shown for a lattice model of finite extension in x direction, modelling the physically more relevant situation of a border instead of an interface where the sign of M changes.

Let us have a closer look at the material system of 2D topological insulator of HgTe/CdTe quantum wells (QW). This system is interesting both from a fundamental point of view, since it is the first system where the QSHE has been detected experimentally, and also from a point of view focussing on applications, since the large and tunable Rashba SO coupling makes it a promising candidate for future spintronic devices, i.e. devices where information processing is done by employing the electron spin. The system can be tuned from the normal to 2D topological insulator phase by changing the thickness d_{QW} of the HgTe layer [Ber+06a; Kön+07]. Conductance measurements in multi-terminal structures [Kön+07; Rot+09] clearly show the existence of one-dimensional helical edge channels in this material for d_{QW} larger than the critical value $d_c = 6.3\,\text{nm}$ [Kön+07; Rot+09].

The well material is HgTe, and the barrier material is an alloy of CdTe with minor admixture of HgTe, usually $(Hg_{0.3}Cd_{0.7})Te$. In the quantum well structures used for the QSHE measurements [Kön+07], extra intrinsic (i.e. undoped) layers of $(Hg_{0.3}Cd_{0.7})Te$ were grown between the HgTe well layer and the doped $(Hg_{0.3}Cd_{0.7})Te$ barrier layers [Kön+08]. This technique is called modulation doping, and the purpose is to reduce scattering of charges in the well at dopant atoms. This allows for very high mobilities ($> 10^5 \frac{\text{cm}^2}{\text{Vs}}$).

Both HgTe and CdTe are II-VI semiconductors and crystallize in the zinc blende structure (each Te atom is surrounded by four Hg/Cd atoms forming the corners of a tetrahedron, the Te sitting in the center, and vice versa for Hg/Cd surrounded by Te). The lattice constants of the constituents are very similar (6.46 Å for HgTe and 6.48 Å for CdTe [Chu+07]). The interested reader can find a compendium about properties of II-VI semiconductors in [Chu+07]. This section is partly based on the comprehensive book [Fra+13] on topological insulators.

In HgTe and CdTe, the bonds are formed by the $6s^2$ electrons of Hg and Cd, and the $5p^4$ electrons of Te. In the zinc blende crystal, the s-orbitals form the bonding Γ_1^+ and the anti-bonding Γ_1^- bands, while the p-orbitals form the bonding Γ_{15}^+ and the anti-bonding Γ_{15}^- bands [Win05]. The Γ_1^+ band is less interesting since it is filled by the $5s^2$ electrons of Te, and the Γ_{15}^- is

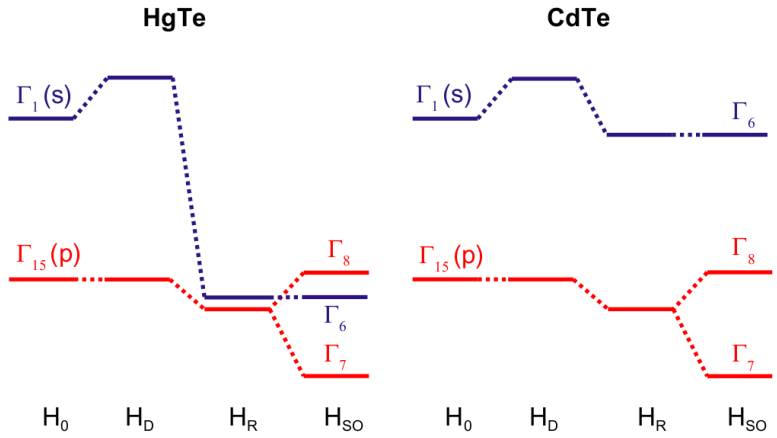


Figure 6.2: A visualization of the influence of the various relativistic corrections on the atomic energy levels of HgTe and CdTe, which form the Γ_6, Γ_7, and Γ_8 bands of the crystal. The interplay of the mass-velocity term H_R and the spin-orbit splitting H_{SO} is responsible for the inverted band structure of HgTe. Figure reprinted with permission from [Fra+13] with adapts [Chu+07]. © (2013) by Elsevier Books

a distant valence band. The interesting bands are the Γ_1^- (valence) and Γ_{15}^+ (conduction) bands, which are filled by the mentioned six outer electrons of HgTe or CdTe. Here we have used the group theoretical notations for the T_d group, not yet caring about SO coupling, compare Fig. 6.2. In presence of SO coupling, the double group representations $\Gamma_6^+(s)$, $\Gamma_7^+(p)$ and $\Gamma_8^+(p)$ are required, compare also the group theoretical discussion in Chapter 3. Energetically, the crystal bands must be found close to the atomic levels. Since we are dealing with heavy atoms, the various relativistic corrections, including SO coupling, are important. After taking these into account, the six outer electrons of an atom pair (Hg/Cd, Te) will have to redistribute among the new levels. The relativistic corrections have been analysed and estimated in [Her+63], and their effect on the energy levels is visualized in Fig. 6.2. The relativistic mass-velocity correction H_R corresponds to order p^4 when expanding the relativistic energy $\sqrt{p^2c^2 + m_0^2c^4}$ in p. The Darwin term H_D and SO coupling term H_{SO} both appear due to the electric field of the core, as effective terms in the non-relativistic limit [Fol+50]. The correction due to the Darwin term is comparable for HgTe and CdTe. In contrast, the corrections due to the mass-velocity term are very different because the masses of Hg and Cd cores are very different, pulling the level of s-orbitals in the heavier Hg much further down. Finally, the SO coupling is responsible for the splitting of the double group representations Γ_8 and Γ_7 that originate from the T_d single group representation Γ_{15} describing the transformation of p-orbitals. Since the p-orbitals originate from Te in both compounds, the splitting is identical. Γ_8 corresponds to the quadruplet of $j = \frac{3}{2}$ states, consisting of the heavy holes (HH) with $m_j = \pm\frac{3}{2}$ and the light holes (LH) with $m_j = \pm\frac{1}{2}$, and Γ_7 corresponds to the SO split-off band with $j = \frac{1}{2}$. In the end, the interplay of mass-velocity corrections and the SO correction is causing the unusual, inverted band structure of HgTe, where Γ_6 lies energetically below Γ_8. Filling up the states with the six electrons per pair of atoms, one finds that only the heavy hole band in HgTe can be populated. At the Γ point ($k = 0$), rotational symmetry enforces fourfold degeneracy, and the HH and LH bands must touch. (Note that this no longer holds in a quasi-2D system like a QW, where rotational

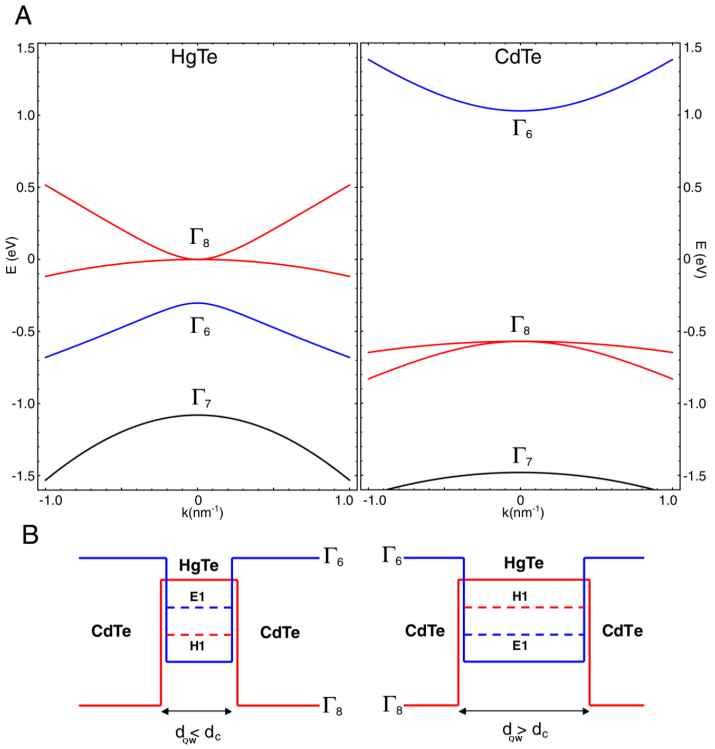


Figure 6.3: (A) Band structure for bulk HgTe and CdTe; (B) Schematic picture of the HgTe/HgCdTe quantum well, shown for a width d_{QW} smaller than the critical width d_c (normal regime) and for $d_{QW} > d_c$ (inverted regime). Figure reprinted with permission from [Kön+08]. © (2008) by the Physical Society of Japan

symmetry is restricted to in-plane rotations.) Due to the band inversion, the LH and HH bands are the conduction and valence bands in HgTe, so this material is a half-metal, meaning that an insulating gap is prevented by symmetry and conduction and valence bands touch. Figure 6.3 shows the band structures of bulk HgTe and bulk CdTe in the vicinity of the Γ point. The band structure of the bulk materials may be obtained in the O_h approximation to T_d, neglecting the bulk inversion asymmetry (BIA) of zinc blende which lifts the twofold spin degeneracy (Kramers' degeneracy) of the bands. Actually, we will be interested in the spin splitting of the QW subbands. But in the QW, the splitting occurs not only due to BIA, but also because of structural inversion asymmetry (SIA), if the QW is grown asymmetrically, or if there is a top or bottom gate which induces an asymmetric charge distribution in the growth direction (which is [001] = z direction). So the SIA can be tuned electrically. Compared to the splitting due to typical (non-zero) SIA values, splitting caused by the BIA is negligible [Nov+05a], which justifies the use of the O_h approximation. Further, the band gap of the 2D TI quantum well system can be chosen large enough that the energy scale of BIA is small in comparison. Then, BIA is not important for the topological classification, which is the reason why it is neglected in [Ber+06a].

In [Nov+05a], calculations for the subband dispersion of HgTe/$(Hg_{0.3}Cd_{0.7})$Te quantum wells are shown. The starting point is the envelope function theory written down for the basis of the

bulk 8×8 Kane Hamiltonian for zinc blende systems [Kan57]. In [Pfe00], the envelope function theory after Burt [Bur88b] (compare Chapter 2) has been applied for a [001] quantum well. From the derivation by Kane [Kan57], it is clear that anisotropy in the spectrum cannot be due to the 8 bands that are included exactly, but appears only due to the coupling to remote bands, which are treated perturbatively (those terms are of 2nd order in k). In the Hamiltonian for the envelope functions of the QW, the rotational symmetry is reduced to discrete rotations about $\hat{z}$. But still, the dispersion remains anisotropic in the plane of $\mathbf{k}_\parallel = (k_x, k_y)$. Matrix elements responsible for the anisotropy can be easily identified since they appear as linear combinations of k_-^2 and k_+^2. In the axial approximation, the in-plane anisotropy is neglected. This is done in [Nov+05a], and we will also do so in our derivation of effective QW subband Hamiltonians in Chapter 7.1.1.

A sketch of the relevant subband levels is given in the lower part of Fig. 6.3. At $k_\parallel = 0$, HH states of the well and barrier do not mix, while LH states and conduction band states mix. The subband edges ($k_\parallel = 0$) can therefore be classified in $|Hn, \pm\rangle$ for the nth heavy hole subband, and admixtures of predominantly electron states $|En, \pm\rangle$ which we will call electron-like, as well as light-hole-like states $|LHn, \pm\rangle$. They are two-fold (Kramers) degenerate. The most interesting QW bands are E_1 and H_1. The LH bands are energetically remote ($E_{LH,n} < -100$ meV).

The different band ordering of HgTe and CdTe competes in the QW subbands. Depending on the QW width d_{QW}, the subband dispersion will have an inverted or normal band structure. This is already clear from a simple consideration of limiting cases. For a wide QW, the H_1 subband will lie above E_1, corresponding to the inverted band structure of HgTe. For a narrow QW, E_1 will be above H_1, taking over the normal band structure of CdTe. Thus, there must be a critical width d_c where E_1 and H_1 cross, if the QW width is varied. The crossing is protected by symmetry, since the states $|E1, \pm\rangle$ are parity-odd and $|H1, \pm\rangle$ are parity-even. The effective 4×4 BHZ Hamiltonian (6.12) captures the physics of the E_1 and H_1 levels in the interesting regime around d_c. There, tuning the width d_{QW} corresponds to tuning the gap parameter M. The block-diagonal form of (6.12) assumes a symmetric QW. The BHZ model adequately describes the insulating regime in HgTe/CdTe QWs close to the Γ point and the topological quantum phase transition near the critical thickness $d_{QW} = d_c$. It has been extended to include the bulk inversion symmetry breaking effects in Ref. [Kön+08]. However, this model does not yet include the structural inversion asymmetry (SIA) terms that can be very large in this narrow gap material. Indeed, it was shown experimentally that an external top gate applied to the HgTe/CdTe QWs can change the energy of the Rashba spin-orbit splitting in the range from 0 to 30meV [Nov+05a] and the samples can be tuned from insulating to metallic regime [Brü+10]. Furthermore, the Aharonov-Casher oscillations [Kön+06] as well as the ballistic spin-Hall effect in HgTe/CdTe QWs [Brü+10], which occur in the metallic regime, can be well described by an effective two band (electron or heavy hole) model taking into account the Rashba spin-orbit interactions. The Aharonov-Casher oscillations can be seen either as generalization of the Aharonov-Bohm oscillations to systems including magnetic field and SO coupling, or as application of the Berry phase to ring structures of such systems. Then, the interference pattern of waves traversing different paths of a ring structure depends on the enclosed flux, SO coupling and spin orientations of the waves.

On the theoretical side, three-dimensional insulators have been predicted and classified by a topological invariant [Fu+07b; Dai+08]. The class of strong topological insulators in 3D possesses surface states which are described by a gapless 2D Dirac cone, and their spectrum is protected in the sense that time-reversal conserving perturbations like disorder cannot open a gap. This

is analogous to the protected gapless 1D edge states of a 2D TI. The simplest way to find topologically nontrivial insulators, is to look at systems where the conduction and valence bands have opposite parity and a change in band ordering (called band inversion) occurs as a function of a tuning parameter like the strength of spin-orbit coupling [Ber+06a; Fu+07a]. This criterion leads to the unified Dirac form of the effective Hamiltonian for topological insulators with a spatial inversion center [Ber+06a; Zha+09]. This method may be applied to 2D as well as 3D systems. A key point is that an insulator is guaranteed to be trivial if there is no SO interaction. In [Zha+09], ab-initio band structure calculations have been carried out for Bi_2Se_3, and by artificially modifying the SO parameter, it is shown how the band inversion occurs, causing Bi_2Se_3 to become a 3D TI with a bulk gap greater than the energy k_BT at room temperature. Angle-resolved photoemission experiments on Bi_2Se_3 thin films [Zha+10] have been carried out to analyse the crossover of 3D to 2D TIs.

Let us also note in passing that the effective Hamiltonian of a 3D TI thin film, in which the overlap of surface states of opposing faces must be considered, looks like (6.12) with parameters replaced [Liu+10a; Lin+09; Lu+10; Sha+10]. Experimentally, SIA is important if the thin film of Bi_2Se_3 is grown on a substrate, but the SIA terms of the surface states [Sha+10] will have a different analytical form than for HgTe/CdTe QWs.

Meanwhile, a full classification scheme of all topological states of band Hamiltonians is available [Sch+09], where the class depends on the spatial dimension of the system and the Cartan-Altland-Zirnbauer class [Alt+97]. The transformation of a system under the anti-unitary operations of time reversal and particle-hole conjugation, and their combined operation (called chiral symmetry operation), is used to define the ten Cartan-Altland-Zirnbauer classes, which appear in the topological classification. Recently, interest has been focussed on the question whether the topological classification is still possible when one allows for more general (many-particle) interactions. This issue is by no means settled. By intuition, it is already clear that adiabatically turning on moderate interactions, which are not strong enough to close the bulk gap, should not change the topological invariant of a non-interacting system. Based on this idea, one can formalize the generalization of topological classification to interacting systems by expressing a topological invariant in terms of the single particle Green's function of the many-particle system [Wan+13].

Part II

Results

Chapter 7

Fingerprint of different spin-orbit terms for spin transport in HgTe quantum wells

In this chapter[1], we show the derivation of an effective four-band model describing the physics of the typical two-dimensional topological insulator (HgTe/CdTe quantum well) in the presence of perturbing potentials. A short introduction to topological insulators has been given in Chapter 6. We include an out-of plane potential that breaks the inversion symmetry in z-direction, causing structural inversion asymmetry (SIA), as well as in-plane potentials. We employ the methods of $\mathbf{k} \cdot \mathbf{p}$ and envelope function theory as described in Chapter 2 and extend the BHZ model (6.12) [Ber+06a]. We show that the result is consistent with the Hamiltonian obtained by the method of invariants (see Chapter 3). The inversion breaking potential generates new elements to the four-band Hamiltonian that are off-diagonal in spin space. The generalized four-band Hamiltonian (7.11) should also be applicable to other 2D topological insulators, such as type II InAs/GaSb quantum wells [Liu+08].

Next, we will use the Foldy-Wouthuysen (FW) transformation to find an effective model describing the electron or heavy hole bands. We show that such an effective model contains two different types of SO interactions; one of them is the well-known Rashba spin-orbit interaction induced by the inversion breaking potential in z-direction, whereas the other originates from the in-plane potential, and is referred to as the in-plane Pauli term. Although both these terms, for the conduction band, are linear in the wave vector, they contribute differently to the spin transport. The first (Rashba) term does not conserve the z-component of spin, S_z, causing spin precession, while the in-plane Pauli term conserves S_z.

Spin transport in the conduction band is further analysed numerically within the Landauer-Büttiker formalism (compare Chapter 5), in order to study the interplay of the Rashba and in-plane Pauli terms. We predict that the spin Hall conductance will show a precession pattern as a function of the inversion breaking potential in the z-direction even in the presence of a strong in-plane potential. Further, the strong in-plane potential enhances the spin Hall conductance generated by the Rashba term, because it partially fixates the direction of the precessing spin. Therefore, the behaviour of the spin transport in asymmetrically doped quantum wells should be dominated by the Rashba term and it is justified to describe the spin Hall conductance in

[1] This chapter is based on the publication [Rot+10].

the metallic regime through simple effective models for electrons and heavy holes, as long as the band gap is non-zero. The spin Hall signal can be qualitatively understood by a spin force operator or a spin-dependent Berry phase appearing in a semiclassical wave packet description (compare Chapter 4).

7.1 Effective Hamiltonian for HgTe QWs in the presence of the inversion breaking potential in z-direction and the in-plane potential

7.1.1 Derivation of the extended HgTe Hamiltonian within k · p theory

In this section we will consider the influence of the structural inversion asymmetry (SIA) on HgTe/CdTe quantum wells (QWs) and derive a corresponding effective 4×4 model with an out-of plane (in z-direction) inversion breaking potential. Our starting point is the envelope function method based on the eight-band Kane Hamiltonian H_{K} [Nov+05a]. The band parameters are z-dependent, and there are interface terms appearing because of non-commuting operators, which vanish for z-independent parameters (compare Section 2.3 about the symmetrization of non-commuting terms). The basis of the Kane Hamiltonian is given by

$$
\begin{aligned}
|1\rangle &= |\Gamma_6^-, 1/2\rangle = |S\rangle\,|\uparrow\rangle\,,\\
|2\rangle &= |\Gamma_6^-, -1/2\rangle = |S\rangle\,|\downarrow\rangle\,,\\
|3\rangle &= |\Gamma_8^+, 3/2\rangle = -\frac{1}{\sqrt{2}}\,(|X\rangle + i|Y\rangle)\,|\uparrow\rangle\,,\\
|4\rangle &= |\Gamma_8^+, 1/2\rangle = \frac{1}{\sqrt{6}}\,(2|Z\rangle\,|\uparrow\rangle - (|X\rangle + i|Y\rangle)\,|\downarrow\rangle)\,,\\
|5\rangle &= |\Gamma_8^+, -1/2\rangle = \frac{1}{\sqrt{6}}\,((|X\rangle - i|Y\rangle)\,|\uparrow\rangle + 2|Z\rangle\,|\downarrow\rangle)\,,\\
|6\rangle &= |\Gamma_8^+, -3/2\rangle = \frac{1}{\sqrt{2}}\,(|X\rangle - i|Y\rangle)\,|\downarrow\rangle\,,\\
|7\rangle &= |\Gamma_7^+, 1/2\rangle = -\frac{1}{\sqrt{3}}\,(|Z\rangle\,|\uparrow\rangle + (|X\rangle + i|Y\rangle)\,|\downarrow\rangle)\,,\\
|8\rangle &= |\Gamma_7^+, -1/2\rangle = \frac{1}{\sqrt{3}}\,(|Z\rangle\,|\downarrow\rangle - (|X\rangle - i|Y\rangle)\,|\uparrow\rangle)\,,
\end{aligned}
\tag{7.1}
$$

where the states $|X\rangle, |Y\rangle, |Z\rangle$ transform like p-orbitals with positive parity and $|S\rangle$ transforms like an s-orbitals and has negative parity. We use the standard notation with $|\Gamma_6^-, \pm1/2\rangle$ describing the s-like conduction band, $|\Gamma_8^+, \pm1/2\rangle$ the p-like light hole band and $|\Gamma_8^+, \pm3/2\rangle$ the p-like heavy hole band in zinc blende crystal structures [Win05]. The spin-orbit split-off bands $|\Gamma_7, \pm1/2\rangle$ are far away in energy from the other bands and are not important for the description of the quantum well, therefore we limit ourself to the upper 6×6 block of H_{K}. Let us emphasize that our 6×6 Kane Hamiltonian preserves the bulk inversion asymmetry, since we omit negligible effects of bulk inversion asymmetry in the zinc blende structure of bulk HgTe and CdTe [Kön+08]. In the following we always use $|\alpha\rangle$ ($\alpha = 1, 2, \ldots, 6$) to denote the basis set of wave functions shortly. We consider a quantum well configuration with HgTe layers sandwiched by two CdTe barrier layers along z direction as in Fig. 6.3, hence the parameters of the Kane model H_{K} have spatial dependence [Ber+06a]. The matching of wave functions in z-direction

for HgTe/CdTe QWs has to be done very carefully because the bulk barrier material CdTe has a normal band structure with $|\Gamma_6^-, \pm 1/2\rangle$ above $|\Gamma_8^+\rangle$ bands while bulk HgTe has inverted band ordering with $|\Gamma_8^+\rangle$ above $|\Gamma_6^-, \pm 1/2\rangle$ bands [Ber+06a; Kön+07].

The envelope function approximation [Bur88b] is applied to solve the eigenproblem of the quantum well. Since the Kane model preserves inversion symmetry, in order to discuss the SIA, we need to take into account an additional potential $V(\mathbf{r}) = V_0(x, y) + ze\mathcal{E}_z$, where $e > 0$ is the elementary charge and $ze\mathcal{E}_z$ is the inversion breaking potential in z-direction, while $V_0(x, y)$ is the in-plane potential and possible forms will be chosen in the numerical transport analysis (Section 7.3). Then the full Hamiltonian is

$$\hat{H}_{\text{full}} = H_{\text{K}}(\mathbf{k}_{\parallel}, z) + V(\mathbf{r}). \tag{7.2}$$

Next we split the Hamiltonian (7.2) into two parts $\hat{H}_{\text{full}} = H_0 + H'$, where H_0 is the Kane Hamiltonian when $k_{\parallel} = 0$ and is treated as the zero-order Hamiltonian. Explicitly, H_0 is given by

$$H_0 = H_{\text{K}}(k_{\parallel} = 0) = \begin{pmatrix} T^{(0)} & 0 & 0 & \sqrt{\frac{2}{3}} P\hat{k}_z & 0 & 0 \\ 0 & T^{(0)} & 0 & 0 & \sqrt{\frac{2}{3}} P\hat{k}_z & 0 \\ 0 & 0 & W_+^{(0)} & 0 & 0 & 0 \\ \sqrt{\frac{2}{3}} P\hat{k}_z & 0 & 0 & W_-^{(0)} & 0 & 0 \\ 0 & \sqrt{\frac{2}{3}} P\hat{k}_z & 0 & 0 & W_-^{(0)} & 0 \\ 0 & 0 & 0 & 0 & 0 & W_+^{(0)} \end{pmatrix} \tag{7.3}$$

where $\hat{k}_z$ is an operator and the heavy hole bands $(\Gamma_8^+, \pm 3/2)$ are completely decoupled from the electron and light hole bands. Here $P = -\frac{\hbar^2}{2m_0}\langle S|p_x|X\rangle$ is the Kane matrix element between the Γ_6 and Γ_8 bands, while the other parameters are given by[2]

$$T^{(0)} = E_c(z) + \frac{\hbar^2}{2m_0}\hat{k}_z(2F(z) + 1)\hat{k}_z \tag{7.4}$$

$$W_{\pm}^{(0)} = E_v(z) - \frac{\hbar^2}{2m_0}\hat{k}_z\left(\gamma_1(z) \mp 2\gamma_2(z)\right)\hat{k}_z \tag{7.5}$$

with $F(z) = \frac{1}{m_0}\sum_j^{\Gamma_5} \frac{|\langle S|p_x|u_j\rangle|^2}{E_c(z) - E_j(z)}$ including remote bands $|u_j\rangle$ with Γ_5 symmetry perturbatively. $E_{c/v}$ designate the positions of the conduction/valence band edges and the γ_i are renormalized Luttinger parameters [Lut56]. The axial approximation is adopted [Eke+85; Pfe00] in order to keep the in-plane rotation symmetry.

[2] Unfortunately, there is a sign mistake in $W_{\pm}^{(0)}$ in our publication [Rot+10]. However, my MATHEMATICA file for calculations based on the 8×8 Kane Hamiltonian in [Pfe00] does not contain this mistake, and therefore subsequent results shown here and in [Rot+10] are not affected by the mistake.

H' is treated as a perturbation, and is written as

$$H' = H_{\mathrm{K}}(\mathbf{k}) - H_0 + V = \begin{pmatrix} T^{(1)} & 0 & -\frac{Pk_+}{\sqrt{2}} & 0 & \frac{Pk_-}{\sqrt{6}} & 0 \\ 0 & T^{(1)} & 0 & -\frac{Pk_+}{\sqrt{6}} & 0 & \frac{Pk_-}{\sqrt{2}} \\ -\frac{Pk_-}{\sqrt{2}} & 0 & W_+^{(1)} & -S_- & R & 0 \\ 0 & -\frac{Pk_-}{\sqrt{6}} & -S_-^\dagger & W_-^{(1)} & C & R \\ \frac{Pk_+}{\sqrt{6}} & 0 & R^\dagger & C^\dagger & W_-^{(1)} & S_+^\dagger \\ 0 & \frac{Pk_+}{\sqrt{2}} & 0 & R^\dagger & S_+ & W_+^{(1)} \end{pmatrix} \tag{7.6}$$

with $k_\pm = k_x \pm ik_y$, $C = \frac{\hbar^2}{2m_0} k_-[\kappa, k_z]$, $R = \frac{\sqrt{3}\hbar^2}{2m_0}\bar{\gamma}k_-^2$, $S_\pm = -\frac{\sqrt{3}\hbar^2}{2m_0}k_\pm(\{\gamma_3, k_z\} + [\kappa, k_z])$, $T^{(1)} = \frac{\hbar^2(2F+1)k_\parallel^2}{2m_0} + V(\mathbf{r})$ and $W_\pm^{(1)} = -\frac{\hbar^2}{2m_0}(\gamma_1 \pm \gamma_2)\, k_\parallel^2 + V(\mathbf{r})$. Here, $\bar{\gamma} = (\gamma_3 + \gamma_2)/2$. κ is the renormalized Luttinger parameter related to the part of Hamiltonian which is antisymmetric in the components of $\mathbf{k}$. In the original Luttinger model, it was introduced because in the presence of a magnetic field the components of $\mathbf{k}$ do not commute. In our case, it appears because the material parameters are functions of the z coordinate.

Now we will generalize the BHZ approach [Ber+06a] to project the Hamiltonian (7.2) into the low energy sub-space, which can be done in two steps. First, we numerically diagonalize the Hamiltonian H_0, so that $H_0|i\rangle = E_i|i\rangle$, to obtain the eigenenergies E_i and eigenstates $|i\rangle$ of the quantum well. Here the eigenstate $|i\rangle$ can be expanded in the basis $|\alpha\rangle$ as $|i\rangle = \sum_\alpha f_{i,\alpha}(z)|\alpha\rangle$, where the function $f_i(z)$ gives the envelope function along z-direction for the quantum well. We use Greek indices to indicate basis functions of the Kane Hamiltonian and Roman indices to denote the subbands. The envelope function components $f_{i,\alpha}(z)$ are calculated with the help of the numerical diagonalization of H_0.

In order to perform the degenerate perturbation calculation (see Appendix B), we need to cast the eigenstates of H_0 into two classes. The first one, denoted as class A, includes the basis wave functions of our final four band effective model. As shown by BHZ [Ber+06a], for HgTe/CdTe quantum wells, it is necessary to take into account the two electron-like subbands $|E1, \pm\rangle$ and two heavy hole subbands $|H1, \pm\rangle$, which are expanded explicitly as

$$|E1,+\rangle = f_{\mathrm{E}+,1}(z)|1\rangle + f_{\mathrm{E}+,4}(z)|4\rangle \tag{7.7a}$$
$$|H1,+\rangle = f_{\mathrm{H}+,3}(z)|3\rangle \tag{7.7b}$$
$$|E1,-\rangle = f_{\mathrm{E}-,2}(z)|2\rangle + f_{\mathrm{E}-,5}(z)|5\rangle \tag{7.7c}$$
$$|H1,-\rangle = f_{\mathrm{H}-,6}(z)|6\rangle \tag{7.7d}$$

As pointed out above, for H_0 the heavy hole bands are decoupled from the electron and light hole bands, therefore the eigenstate $|H1, +(-)\rangle$ consists only of the basis $|3\rangle$ ($|6\rangle$) while $|E1, +(-)\rangle$ is a combination of the basis $|1\rangle$ ($|2\rangle$) and $|4\rangle$ ($|5\rangle$). The second class, denoted as class B, includes the states which need to be taken into account in the following perturbation procedure. Here we consider the first light hole-like subbands $|LH, \pm\rangle$ and the second and third heavy hole subbands $|HH2, \pm\rangle$ and $|HH3, \pm\rangle$, which are written explicitly as

$$|LH,+\rangle = f_{\mathrm{LH}+,1}(z)|1\rangle + f_{\mathrm{LH}+,4}(z)|4\rangle \tag{7.8a}$$
$$|HHn,+\rangle = f_{\mathrm{HHn}+,3}(z)|3\rangle \tag{7.8b}$$
$$|LH,-\rangle = f_{\mathrm{LH}-,2}(z)|2\rangle + f_{\mathrm{LH}-,5}(z)|5\rangle \tag{7.8c}$$
$$|HHn,-\rangle = f_{\mathrm{HHn}-,6}(z)|6\rangle. \tag{7.8d}$$

All the other subbands of the quantum well are neglected here since they are well separated in energy. Due to the structure of H_0, the envelope function components $f_{i,\alpha}(z)$ can be chosen to be either real or purely imaginary.

Before we go to the next step of the perturbation calculation, it is useful to have a look at the symmetry properties of the relevant states. For the Hamiltonian H_0, we have three types of symmetries: the time reversal symmetry $\mathcal{T}$, the inversion symmetry P and the in-plane full rotation symmetry $R_z(\theta)$. For the time reversal operation $\mathcal{T}$, it is not hard to show that $|E1,\pm\rangle$ ($|H1,\pm\rangle$) are Kramers partners, i.e. (the overall sign being convention) $\mathcal{T}|E1,+\rangle = |E1,-\rangle$, $\mathcal{T}|E1,-\rangle = -|E1,+\rangle$, $\mathcal{T}|H1,+\rangle = |H1,-\rangle$ and $\mathcal{T}|H1,-\rangle = -|H1,+\rangle$. When matrix elements for one spin are found, those for the opposite spins can be easily obtained through the operation $\mathcal{T}$, under the precaution that we also flip the sign of $\mathbf{k}_\parallel$. The inversion operation P defines the parity of each subband, which can greatly simplify the matrix elements in the perturbation procedure below. The parity of the subbands $|i\rangle$ in the quantum well is determined by both the envelope function $f_{i,\alpha}(z)$ and the basis wave function $|\alpha\rangle$. The parities of the envelope functions can be obtained through numerical calculation [Ber+06a; Pfe00], and are listed in Table 7.1. The parities of the basis functions are indicated by the parity superscript of the irreducible

even:	$f_{\mathrm{E}+,1}$	$f_{\mathrm{E}-,2}$	$f_{\mathrm{LH}+,4}$	$f_{\mathrm{LH}-,5}$	$f_{\mathrm{H}+,3}$	$f_{\mathrm{H}-,6}$	$f_{\mathrm{HH3}+,3}$	$f_{\mathrm{HH3}-,6}$
odd:	$f_{\mathrm{E}+,4}$	$f_{\mathrm{E}-,5}$	$f_{\mathrm{LH}+,1}$	$f_{\mathrm{LH}-,2}$	$f_{\mathrm{HH2}+,3}$	$f_{\mathrm{HH2}-,6}$		

Table 7.1: Parities of the envelope function components.

representation. Thus the parities of the subbands are $P|E1\pm\rangle = -|E1\pm\rangle$, $P|H1\pm\rangle = |H1\pm\rangle$, $P|LH\pm\rangle = |LH\pm\rangle$, $P|HH2\pm\rangle = -|HH2\pm\rangle$ and $P|HH3\pm\rangle = |HH3\pm\rangle$. Due to the in-plane rotation symmetry (axial approximation), the total angular momentum J along z-direction is a good quantum number, which can be used to identify the eigenstates. Since the electron-like subbands have $J = \frac{1}{2}$, the rotation operator is $R_z(\theta)|E1\pm\rangle = e^{\pm i\frac{\theta}{2}}|E1\pm\rangle$ while for the heavy hole subbands with $J = \frac{3}{2}$, it should be $R_z(\theta)|H1\pm\rangle = e^{\pm i\frac{3\theta}{2}}|H1\pm\rangle$.

Next, we calculate the effective Hamiltonian of the four states in the class A based on quasi-degenerate perturbation theory (see Appendix B or [Win05]). All states in classes A and B are eigenstates of Hamiltonian H_0. However when H' is introduced, they are no longer eigenstates due to mixing between the states of class A and class B. Therefore, treating H' as a small perturbation, we need to perform an unitary transformation to eliminate the coupling between the states in class A and class B up to the required order, resulting in the transformed Hamiltonian $\tilde{H}$. We apply the third order perturbation formula (B.12) to find matrix elements of $\tilde{H}$ in terms of matrix elements of H' obtained by integrations over the QW growth direction, like

$$H'_{jk} := \langle f_j|H'|f_k\rangle = \int dz \sum_{\alpha,\beta=1}^{6} f^*_{j,\alpha}(z)(H')_{\alpha\beta}\, f_{k,\beta}(z). \tag{7.9}$$

This is different from the usual procedure of perturbation theory, where one would first evaluate the perturbation series and then carry out the integration over z, to obtain an effective Hamiltonian. Since we use a finite (i.e. not a complete) set of envelope functions, performing the integration first, like here, amounts to an additional approximation. As mentioned above, Greek indices denote entries of the Kane matrix (7.6). Here we should keep in mind that the order of the matrix elements of H' in (B.12) is important, as they may not commute with each other.

The perturbation calculation is straightforward but lengthy. The parities of the envelope functions discussed above can be used to reduce the number of the matrix elements of H'. For example, the first-order term

$$\langle E1-|H'|E1+\rangle = -\frac{Pk_-}{\sqrt{6}}\langle f_{\mathrm{E}-,2}|f_{\mathrm{E}+,4}\rangle + \frac{Pk_-}{\sqrt{6}}\langle f_{\mathrm{E}-,5}|f_{\mathrm{E}+,1}\rangle + \langle f_{\mathrm{E}-,5}|C^\dagger|f_{\mathrm{E}+,4}\rangle \tag{7.10}$$

vanishes completely because all integrands are odd functions of z (see Table 7.1).

By applying time reversal to (7.7) and (7.8), we find the relations $f_{\mathrm{E}-,2} = f^*_{\mathrm{E}+,1}$, $f_{\mathrm{E}-,5} = -f^*_{\mathrm{E}+,4}$, $f_{\mathrm{H}-,6} = f^*_{\mathrm{H}+,3}$, $f_{\mathrm{LH}-,2} = -f^*_{\mathrm{LH}+,1}$, $f_{\mathrm{LH}-,5} = f^*_{\mathrm{LH}+,4}$ and $f_{\mathrm{HH2}-,6} = -f^*_{\mathrm{HH2}+,3}$. These relations will be used to write resulting matrix elements using only the + subset of envelope functions. This leads to some simplifications, and also has the advantage that the complex phase of resulting matrix elements can be analysed easily, since time reversal symmetry will no longer enforce any restricting condition on them.

In the four-band basis $(|E1+\rangle, |H1+\rangle, |E1-\rangle, |H1-\rangle)$, the final effective Hamiltonian is written as

$$\tilde{H} = \tilde{H}_0 + \tilde{H}_\mathrm{R} + V_0(x,y) \tag{7.11}$$

$$\tilde{H}_0 = \epsilon(k)I + \begin{pmatrix} \mathcal{M}(k) & \mathcal{A}k_+ & 0 & 0 \\ \mathcal{A}k_- & -\mathcal{M}(k) & 0 & 0 \\ 0 & 0 & \mathcal{M}(k) & -\mathcal{A}k_- \\ 0 & 0 & -\mathcal{A}k_+ & -\mathcal{M}(k) \end{pmatrix} \tag{7.12}$$

$$\tilde{H}_\mathrm{R} = \begin{pmatrix} 0 & 0 & -iR_0k_- & -iS_0k_-^2 \\ 0 & 0 & iS_0k_-^2 & iT_0k_-^3 \\ iR_0k_+ & -iS_0k_+^2 & 0 & 0 \\ iS_0k_+^2 & -iT_0k_+^3 & 0 & 0 \end{pmatrix} \tag{7.13}$$

with $\mathcal{A} = A + A_2k^2$, $\mathcal{M}(k) = M - Bk^2$, $\epsilon(k) = C - Dk^2$, and I is diagonal unit matrix. Here, we have set $k^2 = k_x^2 + k_y^2$.

We note that $\tilde{H}_0$ is equivalent to the BHZ Hamiltonian in [Ber+06a] if we further omit the k-dependence of $\mathcal{A}$, by setting $A_2 = 0$. We also assume that the reference energy is fixed in the middle of the gap, i.e. $C = 0$. Besides the BHZ Hamiltonian, we find a new term $\tilde{H}_\mathrm{R}$, which is off-diagonal in spin space due to the inversion breaking potential $ze\mathcal{E}_z$. As mentioned above we have included the subbands $|HH3,\pm\rangle$ in the calculation of the effective parameters A,B,D. However, these subbands do not contribute to $\tilde{H}_\mathrm{R}$. This is a consequence of the fact that the envelope functions belonging to $|HH3,\pm\rangle$ have parities opposite to the envelope functions of $|HH2,\pm\rangle$, see Table 7.1.

There are three new terms in $\tilde{H}_\mathrm{R}$. The first term (R_0 term) originates from the second order perturbation theory and is exactly the electron Rashba term with

$$\begin{aligned} R_0 = &\frac{ie\mathcal{E}_z}{3(E_\mathrm{LH} - E_\mathrm{E1})}\Bigg[(\langle f_{\mathrm{E}+,1}|z|f_{\mathrm{LH}+,1}\rangle + \langle f_{\mathrm{E}+,4}|z|f_{\mathrm{LH}+,4}\rangle) \\ &\cdot \left(\sqrt{6}P\langle f_{\mathrm{LH}+,4}|f^*_{\mathrm{E}+,1}\rangle + \sqrt{6}P\langle f_{\mathrm{LH}+,1}|f^*_{\mathrm{E}+,4}\rangle + \frac{3\hbar^2}{m_0}\langle f_{\mathrm{LH}+,4}|[\kappa,k_z]|f^*_{\mathrm{E}+,4}\rangle \right) \Bigg]. \end{aligned} \tag{7.14}$$

Here $\langle f_{i,\alpha}|O|f_{j,\beta}\rangle = \int dz f^*_{i,\alpha}(z) O f_{j,\beta}(z)$ for an arbitrary operator O. The electron Rashba term is linear in k because of the $\frac{1}{2}$ electron spin. Conservation of total (orbital plus spin) angular

momentum allows the spin flip operator $\sigma_\pm$ to be combined only with $k_\mp$, it is $[J_z, \sigma_+ k_-] = \hbar[xk_y - yk_x + \frac{1}{2}\sigma_z, \sigma_+ k_-] = 0$.

The second term (T_0 term) originates from the third order perturbation and denotes the heavy-hole k^3 Rashba term with the parameter

$$T_0 = \frac{-i\sqrt{3}e\mathcal{E}_z\hbar^2}{8m_0^2} \cdot (s_1 + s_2 + s_3) \tag{7.15}$$

with

$$\begin{aligned} s_1 =& \frac{2}{(E_{\mathrm{E1}} - E_{\mathrm{HH2}})(E_{\mathrm{H1}} - E_{\mathrm{HH2}})} \Big[\langle f_{\mathrm{E}+,4} | \bar{\gamma} | f^*_{\mathrm{HH2}+,3} \rangle \langle f^*_{\mathrm{HH2}+,3} | z | f^*_{\mathrm{H}+,3} \rangle \\ & \cdot \left(\sqrt{2} P m_0 \langle f_{\mathrm{H}+,3} | f_{\mathrm{E}+,1} \rangle - \sqrt{3}\hbar^2 \langle f_{\mathrm{H}+,3} | (\{\gamma_3, k_z\} + [\kappa, k_z]) | f_{\mathrm{E}+,4} \rangle \right) \Big], \end{aligned} \tag{7.16a}$$

$$\begin{aligned} s_2 =& \frac{2}{(E_{\mathrm{E1}} - E_{\mathrm{LH}})(E_{\mathrm{H1}} - E_{\mathrm{LH}})} \Big[\left(\langle f_{\mathrm{E}+,1} | z | f_{\mathrm{LH}+,1} \rangle + \langle f_{\mathrm{E}+,4} | z | f_{\mathrm{LH}+,4} \rangle \right) \langle f_{\mathrm{LH}+,4} | \bar{\gamma} | f^*_{\mathrm{H}+,3} \rangle \\ & \cdot \left(\sqrt{2} P m_0 \langle f_{\mathrm{H}+,3} | f_{\mathrm{E}+,1} \rangle - \sqrt{3}\hbar^2 \langle f_{\mathrm{H}+,3} | (\{\gamma_3, k_z\} + [\kappa, k_z]) | f_{\mathrm{E}+,4} \rangle \right) \Big], \end{aligned} \tag{7.16b}$$

$$\begin{aligned} s_3 =& \frac{4}{(E_{\mathrm{H1}} - E_{\mathrm{HH2}})(E_{\mathrm{H1}} - E_{\mathrm{LH}})} \Big[\langle f_{\mathrm{LH}+,4} | \bar{\gamma} | f^*_{\mathrm{H}+,3} \rangle \langle f_{\mathrm{H}+,3} | z | f_{\mathrm{HH2}+,3} \rangle \\ & \cdot \left(\sqrt{2} P m_0 \langle f_{\mathrm{HH2}+,3} | f_{\mathrm{LH}+,1} \rangle - \sqrt{3}\hbar^2 \langle f_{\mathrm{HH2}+,3} | (\{\gamma_3, k_z\} + [\kappa, k_z]) | f_{\mathrm{LH}+,4} \rangle \right) \Big]. \end{aligned} \tag{7.16c}$$

Since for heavy holes the spin is $\frac{3}{2}$, the change of angular momentum upon a spin-flip is 3, which corresponds to $k_\pm^3$.

The third term (S_0 term), which is proportional to k^2, also comes from the second order perturbation with the parameter

$$\begin{aligned} S_0 =& -\frac{\sqrt{3}\hbar^2 e\mathcal{E}_z}{4m_0} \Big[\left(\frac{1}{E_{\mathrm{E1}} - E_{\mathrm{HH2}}} + \frac{1}{E_{\mathrm{H1}} - E_{\mathrm{HH2}}} \right) \langle f_{\mathrm{H}+,3} | z | f_{\mathrm{HH2}+,3} \rangle \langle f_{\mathrm{HH2}+,3} | \bar{\gamma} | f^*_{\mathrm{E}+,4} \rangle \\ & + \left(\frac{1}{E_{\mathrm{E1}} - E_{\mathrm{LH}}} + \frac{1}{E_{\mathrm{H1}} - E_{\mathrm{LH}}} \right) \langle f_{\mathrm{H}+,3} | \bar{\gamma} | f^*_{\mathrm{LH}+,4} \rangle \left(\langle f^*_{\mathrm{LH}+,1} | z | f^*_{\mathrm{E}+,1} \rangle + \langle f^*_{\mathrm{LH}+,4} | z | f^*_{\mathrm{E}+,4} \rangle \right) \Big]. \end{aligned} \tag{7.17}$$

This is a new off-diagonal term between the electron-like $\frac{1}{2}$ ($-\frac{1}{2}$) and heavy-hole $-\frac{3}{2}$ ($\frac{3}{2}$) states and the change of the angular momentum is 2, corresponding to $k_\pm^2$. All parameters can be determined by using the numerically calculated $f_{i,\alpha}(z)$, and are listed in the Table 7.2 for a QW with a thickness $d_{QW} = 70$Å.

As we see, all parameter values are real. Interestingly, this is not mere luck, but an effect of a mirror symmetry of the system, which we have not yet mentioned. The QW system within the axial approximation, but including SIA terms, has the point group $C_{\infty v}$. The v subscript indicates that there is a group element $\hat{C}_v$ acting as mirror operation at a plane including the z-axis [Rot+14]. Let us set $V_0(x, y) = 0$ for the moment, since this perturbation could break

A [nm eV]	0.365
B [nm^2 eV]	-0.706
D [nm^2 eV]	-0.532
M [eV]	$-10.09 \cdot 10^{-3}$
$R_0/(e\mathcal{E}_z)$ [nm^2]	-15.6
$T_0/(e\mathcal{E}_z)$ [nm^4]	-8.91
$S_0/(e\mathcal{E}_z)$ [nm^3]	-2.10

Table 7.2: Parameters of the effective 4×4 Hamiltonian, calculated for the quantum well width 70 Å, at charge density $n = 2 \cdot 10^{10}$ cm^{-2}

the mirror symmetry. With the representation $\hat{C}_v(k_x, k_y)\hat{C}_v^{-1} = (-k_x, k_y)$, $\hat{C}_v|E\pm\rangle = |E\mp\rangle$, $\hat{C}_v|H\pm\rangle = |H\mp\rangle$, we find that $[\tilde{H}, \hat{C}_v] = 0$. This forces the parameters $\mathcal{A}$, R_0, S_0 and T_0 to be real. As mentioned above, this is not yet enforced by time reversal symmetry alone. Since the SIA matrix elements (7.14), (7.15) and (7.17) have been written using only the + subspace of Kramers partners, time reversal symmetry imposes no further restriction to the complex phases of these matrix elements. By inspection, we see that there are only two degrees of freedom, given by the complex phases of $|f_{\mathrm{H}+,3}\rangle$ and $|f_{\mathrm{E}+,4}\rangle$, for the three independent SIA matrix elements. However, these two phases are also fixed, once a convention is adopted for the operator $\hat{C}_v$. In this sense, the C_v symmetry fixes one parameter phase and the two others are fixed by convention. By a convention, we mean that unitary transformations which would change the complex phases of the parameters R_0, S_0, T_0 or $\mathcal{A}$, but would also cause $\mathcal{T}$ or $\hat{C}_v$ to transform non-trivially, should be discarded. Stated differently, changing the complex phases of the parameters in an arbitrary way does not always correspond to a unitary transformation, and may instead change the spectrum.

We note that all three terms in the inversion breaking Hamiltonian $\tilde{H}_R$ are proportional to $\mathcal{E}_z$ but are independent of $V_0(x, y)$. The corrections originating from $V_0(x, y)$ to the 2×2 off-diagonal blocks in (7.13) are of higher order in the perturbation than the ones coming from $\mathcal{E}_z$. Moreover, they must contain both $\mathcal{E}_\parallel = \nabla V_0(x, y)/e$ and $\mathcal{E}_z$ due to the fact that the in-plane field $\mathcal{E}_\parallel$ does not break the $z \to -z$ inversion symmetry. Furthermore, $V_0(x, y)$ introduces corrections of third or higher order to the element $(\tilde{H}_0)_{12}$ which have the form $[k_+, [k_+, V_0(x, y)]]k_-$ and $\nabla\mathcal{E}_\parallel \cdot k_+$. These corrections are much smaller than the element Ak_+ which appears already in the first order of perturbation theory in $\tilde{H}_0$. Corrections to the diagonal elements of $\tilde{H}_0$ induced by the in-plane potential are also very small. The latter corrections, after folding to the electron or heavy hole subbands (see Section 7.2), produce similar contributions to the ones originating from $\tilde{H}_0$, but are an order of magnitude smaller due to a large energy separation between main bands and bands which are treated perturbatively. Therefore, the only significant contribution to $\tilde{H}$ connected with the in-plane potential comes from the bare diagonal potential $V_0(x, y)$ as shown in (7.11).

7.1.2 Symmetry arguments for the validity of the extended HgTe Hamiltonian.

The goal of this subsection is to derive the effective 4×4 Hamiltonian (7.11) using the theory of invariants (see Section 3). We remind ourselves that the theory of invariants states that

the Hamiltonian must be invariant under all symmetry operations of the considered system. As mentioned in the last section, the system has time reversal symmetry $\mathcal{T}$, space inversion symmetry P and in-plane rotation symmetry $R_z(\theta)$, and also a reflection symmetry $\hat{C}_v$ at an axis including $\hat{z}$. The transformation of the set of basis wave functions under these symmetries for the effective model has been discussed in the last section. The symmetry operations in the matrix form for the basis $(|E1+\rangle, |H1+\rangle, |E1-\rangle, |H1-\rangle)$ are given by

- *Time reversal symmetry*: $\mathcal{T} = \Theta K$, where $\Theta = -i\sigma_2 \otimes 1$ and K is the complex conjugate operator.
- *Inversion symmetry*: $P = -1 \otimes \tau_3$, the system being even under inversion except for the SIA terms
- *Rotation symmetry*: $R_z(\theta) = e^{i\frac{\Sigma_z}{2}\theta}$ with $\Sigma_z = \sigma_3 \otimes (\frac{1+\tau_3}{2} + \frac{3(1-\tau_3)}{2}) = \sigma_3 \otimes (2 - \tau_3)$
- *Reflection symmetry*: $C_v = \sigma_x \otimes 1$ and $\hat{C}_v(k_x, k_y)\hat{C}_v^{-1} = (-k_x, k_y)$

where the σ_i denote Pauli matrices acting on the spin basis and the τ_i represent Pauli matrices acting on the electron or heavy-hole subbands.

Rather than applying the method of invariants in the standard way as outlined in Chapter 3, we will proceed a bit differently by finding basis matrices from an expansion in a Clifford algebra, and not by the Wigner-Eckart theorem. This leads to more elegant and simpler calculations, while the relation to the method of invariants remains clear. The simplifications are possible because we are dealing with rotation symmetry restricted to the z axis.

A Clifford algebra [Sny97] is the algebra generated by a set of linear independent and anti-commuting matrices. In our case, we need a basis for the 16-dimensional real vector space of Hermitian 4×4 matrices. The full Clifford algebra obtained by the 5 generators $\Gamma_1, ..., \Gamma_5$ will be the $2^5 = 32$-dimensional vector space of general complex matrices, but we need only the Hermitian subspace. The generators anti-commute, $\{\Gamma_i, \Gamma_j\} = 2\delta_{ij}I$. Therefore, products of generators can always be brought into normal ordering. Counting all possible $\begin{pmatrix} 5 \\ k \end{pmatrix}$ products of k different generators, we see that algebra has dimension $\sum_{k=0}^{5} \begin{pmatrix} 5 \\ k \end{pmatrix} = 2^5$. To be specific, we define the generators

$$\Gamma_1 = \sigma_1 \otimes \tau_1, \quad \Gamma_2 = \sigma_2 \otimes \tau_1, \quad \Gamma_3 = \sigma_3 \otimes \tau_1, \quad \Gamma_4 = 1 \otimes \tau_2, \quad \Gamma_5 = 1 \otimes \tau_3, \tag{7.18}$$

and ten commutators of Γ matrices $\Gamma_{ij} = \frac{1}{2i}[\Gamma_i, \Gamma_j]$,

$$\Gamma_{ij} = \varepsilon_{ijk}\sigma_k \otimes 1, \tag{7.19a}$$
$$\Gamma_{i4} = \sigma_i \otimes \tau_3, \quad \Gamma_{i5} = -\sigma_i \otimes \tau_2, \tag{7.19b}$$
$$\Gamma_{45} = 1 \otimes \tau_1 \tag{7.19c}$$

where $i, j = 1, 2, 3$. We have to classify these 16 (including the identity) Hermitian basis matrices according to their transformation under time reversal, parity and rotations about the z axis. Generally, any 4×4 Hamiltonian can be expanded using Γ matrices as

$$\hat{H}_{\text{eff}} = \epsilon(\mathbf{k})I + \sum_i d_i(\mathbf{k})\Gamma_i + \sum_{ij} d_{ij}(\mathbf{k})\Gamma_{ij} \tag{7.20}$$

where I is the 4×4 identity matrix. $\epsilon(\mathbf{k})$, $d_i(\mathbf{k})$ and $d_{ij}(\mathbf{k})$ can be expanded as polynomials of the momentum $\mathbf{k} = (k_x, k_y)$. The Hamiltonian should be invariant under the symmetry operations P, $\mathcal{T}$ and R_z, which indicates that $d_{i(j)}(\mathbf{k})$ should behave the same as $\Gamma_{i(j)}$. Therefore, we need to work out the transformation of the $d_{i(j)}(\mathbf{k})$ and the Γ matrices.

For the fifteen Γ matrices, it is easy to calculate the symmetry transformation under the time reversal operation $\mathcal{T}$ and inversion operation P, which are listed in Table 7.3. For the in-plane rotation operation $R_z(\theta)$, we can calculate the transformation rule $\Gamma'(\theta) = e^{i\frac{\Sigma_z}{2}\theta}\Gamma e^{-i\frac{\Sigma_z}{2}\theta}$ with the help of the differential equation

$$\frac{d\Gamma'(\theta)}{d\theta} = \frac{i}{2}[\Sigma_z, \Gamma'(\theta)]. \tag{7.21}$$

For example, we find that $[\Sigma_z, \Gamma_1] = 4i\Gamma_2$ and $[\Sigma_z, \Gamma_2] = -4i\Gamma_1$, with the solution given by

$$\Gamma_1'(\theta) = \Gamma_1\cos(2\theta) - \Gamma_2\sin(2\theta) \qquad \Gamma_2'(\theta) = \Gamma_1\sin(2\theta) + \Gamma_2\cos(2\theta). \tag{7.22}$$

Thus, (Γ_1, Γ_2) transforms like a vector under rotations. The obtained results are summarized in Table 7.3. We see that Γ_5, Γ_{12} and Γ_{34} behave as a scalar under the rotation R_z; (Γ_4, Γ_3), $(\Gamma_{45}, \Gamma_{35})$ and $(\Gamma_{14}+\Gamma_{23}, \Gamma_{31}+\Gamma_{24})$ rotate as a vector with angular momentum 1; (Γ_1, Γ_2) and $(\Gamma_{15}, \Gamma_{25})$ correspond to angular momentum 2; and $(\Gamma_{23}-\Gamma_{14}, \Gamma_{31}-\Gamma_{24})$ corresponds to angular momentum 3. In Table 7.3, we also list the corresponding tensors formed by $\mathbf{k}$ up to the order k^3. The constructions of these basis polynomials is also done a bit differently than in Chapter 3. We could use the general method, and find the basis by looking at the decomposition of a Cartesian tensor k_ik_j into irreducible parts, but since we here consider only rotations about the z axis, it is much simpler to expand products $k_+^n k_-^m$ and sort by real and imaginary parts.

	R_z	$\mathcal{T}$	P
$\mathbf{I}$	0	+	+
(Γ_1, Γ_2)	2	-	-
(Γ_4, Γ_3)	1	-	-
$(\Gamma_{15}, \Gamma_{25})$	2	+	-
$(\Gamma_{45}, \Gamma_{35})$	1	+	-
Γ_5	0	+	+
Γ_{12}	0	-	+
Γ_{34}	0	-	+
$(\Gamma_{14}+\Gamma_{23}, \Gamma_{31}+\Gamma_{24})$	1	-	+
$(\Gamma_{23}-\Gamma_{14}, \Gamma_{31}-\Gamma_{24})$	3	-	+
(k_x, k_y)	1	-	-
$k_x^2+k_y^2$	0	+	+
$(k_x^2-k_y^2, 2k_xk_y)$	2	+	+
$(k_x^3-3k_xk_y^2, 3k_x^2k_y-k_y^3)$	3	-	-
$(k_x^3+k_xk_y^2, k_x^2k_y+k_y^3)$	1	-	-

Table 7.3: Summary of the symmetry properties of Γ matrices and the tensors formed by $\mathbf{k}$.

From Table 7.3, if we want to preserve $\mathcal{T}$, P and rotation symmetry, then up to k^3 the general form of the Hamiltonian is given by

$$\tilde{H}_0 = \epsilon_k + \mathcal{M}(k)\Gamma_5 + \mathcal{A}(k)(\Gamma_4, \Gamma_3) \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} k_x \\ k_y \end{pmatrix}, \tag{7.23}$$

where $\mathcal{A}(k) = A + A_2 k^2$, $\mathcal{M}(k) = M - Bk^2$ and the phase θ represents the relative phase between $|E1+\rangle$ and $|H1+\rangle$, which can be chosen arbitrarily. Taking $\theta = -\frac{\pi}{2}$, in (7.23), we recover the BHZ Hamiltonian (7.12) [Ber+06a].

We now consider additional terms which preserve rotation symmetry and time reversal symmetry, but break the inversion symmetry. By inspecting Table (7.3), the following three terms are possible,

$$\begin{aligned} \tilde{H}_R =& \frac{\mathcal{R}(k)}{2}(\Gamma_{14} + \Gamma_{23}, \Gamma_{31} + \Gamma_{24}) \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} k_x \\ k_y \end{pmatrix} \\ &+ \frac{T_0}{2}(\Gamma_{23} - \Gamma_{14}, \Gamma_{31} - \Gamma_{24}) \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} k_x^3 - 3k_x k_y^2 \\ 3k_x^2 k_y - k_y^3 \end{pmatrix} \\ &+ S_0(\Gamma_{15}, \Gamma_{25}) \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} k_x^2 - k_y^2 \\ 2k_x k_y \end{pmatrix}, \end{aligned} \tag{7.24}$$

where $\mathcal{R}(k) = R_0 + R_2 k^2$. Similar to θ, two of the phase factors ϕ, φ and ψ are also arbitrary. However, the third phase is not arbitrary since not all the summands in (7.24) commute, and its choice can affect the energy spectrum.[3] The effect of the 2×2 rotation matrices appearing in (7.24) can be analysed easily noting that it corresponds to substituting $k_\pm = ke^{\pm i\phi_k} \to ke^{\pm i(\phi_k - \phi - 3\psi - 2\varphi)}$. The phase angles can be changed by unitary transformations, but not all of them independently. The physical origin of this subtle point lies in the reflection symmetry $\hat{C}_v$, which we have not yet included in our symmetry-based analysis. As already mentioned in the previous section, the representation of the operation $\hat{C}_v$ is also arbitrary to some degree, since it also transforms under unitary basis transformations. The chosen form of the matrix $C_v = \sigma_x \otimes 1$ fixes two phase angles. Demanding that $[\tilde{H}, \hat{C}_v] = 0$, fixes the third phase angle. If we choose $\theta = -\frac{\pi}{2}$, $\phi = -\frac{\pi}{2}$, $\psi = \frac{\pi}{2}$ and $\varphi = \frac{\pi}{2}$, then the Hamiltonian (7.11) is recovered. This shows that the derivation in Section 7.1.1 has yielded the Hamiltonian with the correct structure, in accordance with all the symmetries of the system.

7.2 Foldy-Wouthuysen transformation of the effective HgTe Hamiltonian

The goal of this section is to obtain an effective 2×2 Hamiltonian (where 2 stands for the spin degree of freedom) for electron $|E1\pm\rangle$ and heavy hole $|H1\pm\rangle$ subbands including non-zero in-plane and out-of plane electric fields. So far, the quantum Spin Hall effect (QSHE) was described by the Hamiltonian used in [Ber+06a], where only the diagonal blocks of our Hamiltonian (7.11) were taken to be non-zero, i.e. for $\mathcal{E}_z, \mathcal{E}_\parallel = 0$. Such a block-diagonal Hamiltonian of a HgTe QW is isomorphic to the Dirac Hamiltonian describing the relativistic motion of an electron in two dimensions ($\hat{p}_z = 0$), which couples particle and antiparticle components with

[3] Unfortunately, we have overlooked this point in our publication [Rot+10] since we did not consider the effect of reflection symmetry, but the results obtained by the $\mathbf{k} \cdot \mathbf{p}$ calculation correctly include that symmetry.

the same spin direction. Here, we start from the full Hamiltonian (7.11) and consider the low energy physics with the energy scale smaller than the gap $2M$. In this case, we can apply the perturbation formula (B.12) to obtain an effective model for electron and hole subbands. This procedure is equivalent to the Foldy-Wouthuysen (FW) transformation [Fol+50], which reduces the relativistic Dirac equation in a potential to the Pauli equation [Win05] (see also Chapter 4.2). We keep terms up to linear order in the in-plane $\mathcal{E}_\parallel$ and out-of plane $\mathcal{E}_z$ electric fields, as well as the terms up to the third order in k. Then the effective Hamiltonians for electron ($\hat{H}_e$) and hole ($\hat{H}_h$) subbands are given by

$$\begin{aligned}\hat{H}_e =& M + V_0(x,y) + \left(-D - B + \frac{A^2}{2M}\right)k^2 + \frac{A^2}{8M^2}e\nabla\mathcal{E}_\parallel \\ & - \mathcal{R}(k)(\boldsymbol{\sigma}\times\mathbf{k})_z + \mathcal{G}(k)(e\mathcal{E}_\parallel\times\mathbf{k})_z\sigma_z \qquad (7.25)\\ \hat{H}_h =& - M + V_0(x,y) + \left(B - D - \frac{A^2}{2M}\right)k^2 + \frac{1}{2}\left(\mathcal{Q}(k)\sigma_+k_-^3 + \mathcal{Q}(k)^\dagger\sigma_-k_+^3\right) \\ & + \frac{A^2}{8M^2}e\nabla\mathcal{E}_\parallel + \mathcal{G}(k)(e\mathcal{E}_\parallel\times\mathbf{k})_z\sigma_z + \frac{1}{2}\left(\frac{AS_0}{2M^2}[k_-,[k_-,V_0]]\sigma_+k_- + h.c.\right) \qquad (7.26)\end{aligned}$$

with

$$\mathcal{G}(k) = \frac{A^2}{4M^2} \qquad (7.27)$$

$$\mathcal{R}(k) = R_0 + \left(\frac{iAS_0}{M} - \frac{A^2}{4M^2}R_0\right)k^2 + \frac{iAS_0}{2M^2}e\nabla\mathcal{E}_\parallel \qquad (7.28)$$

$$\mathcal{Q}(k) = iT_0 + \frac{AS_0}{M} + \frac{iA^2R_0}{4M^2}. \qquad (7.29)$$

The spin-dependent term $\mathcal{G}(e\mathcal{E}_\parallel\times\mathbf{k})_z\sigma_z$ and the spin-independent term $\frac{A^2}{8M^2}e\nabla\mathcal{E}_\parallel$ originate directly from the FW transformation from the Dirac type Hamiltonian in the external potential $\tilde{H}_0 + V_0(x,y)$ (see Eq. (7.11)) to a Pauli type equation. Therefore, by analogy to the relativistic electron in vacuum, we call $\mathcal{G}(e\mathcal{E}_\parallel\times\mathbf{k})_z\sigma_z$ the in-plane Pauli term, while the term $\frac{A^2}{8M^2}e\nabla\mathcal{E}_\parallel$ we call the Darwin term. Note that the Pauli term can be also visualized as resulting from a Rashba field due to the edges of a typical mesa structure used in experiments or as coming from the atomic spin-orbit splitting, but it is only active at the edges where $\mathcal{E}_\parallel$ is finite. The Darwin term does not include a contribution from the field in z-direction due to the assumption that $\mathcal{E}_z$ is constant. The in-plane Pauli and Darwin terms appear both in the electron and hole effective Hamiltonians. The additional terms which are proportional to $\mathcal{R}(k)$ and $\mathcal{Q}(k)$ originate from $\tilde{H}_\mathrm{R}$ and are direct consequence of the broken space inversion symmetry in the z-direction. These terms are usually called Rashba terms and they give linear and cubic in k contributions for electron and heavy hole subsystems, correspondingly. In a typical experimental setup, Rashba terms are generated by an asymmetric doping profile surrounding the quantum well and can be adjusted by a top-gate which induces a tunable electric field in z-direction. Figures 7.1a,b show the magnitude $\frac{A^2}{4M^2}$ of the in-plane spin-orbit interaction (SOI) and electron Rashba coefficient $R_0/(e\mathcal{E}_z)$ as a function of the thickness d_{QW} of the HgTe/CdTe QW. Note that the coupling strength $\frac{A^2}{4M^2}$ for the Pauli term decreases with d_{QW} while the strength of Rashba coupling $R_0/(e\mathcal{E}_z)$ increases with d_{QW}. The origin of the different behaviours of these two SOIs can be understood from the plot of energy versus d_{QW} (see Fig. 7.1c). The in-plane term $\frac{A^2}{4M^2}$ comes from the coupling between the electron and the heavy hole $|H1\pm\rangle$ subbands, and the energy

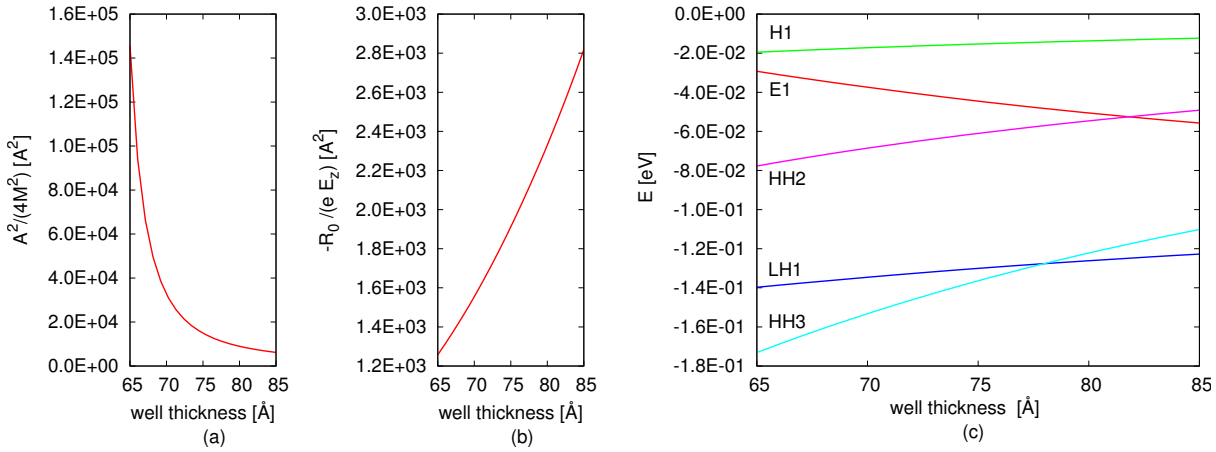


Figure 7.1: The material parameters characterising the spin-orbit coupling induced by (a) the in-plane field ($\frac{A^2}{4M^2}$), (b) the out-of plane field ($R_0/(e\mathcal{E}_z)$) as a function of the quantum well thickness d_{QW}, (c) energies of the most relevant subbands as a function of the QW width for the Γ point. Taken from [Rot+10]. © (2010) Creative Commons CC BY-NC-SA

difference between these bands increases with d_{QW}. The Rashba term $R_0/(e\mathcal{E}_z)$ originates from the coupling between the electron and the light hole sub-bands. Their energy difference decreases with d_{QW}, therefore $R_0/(e\mathcal{E}_z)$ increases with d_{QW}. Comparing the magnitudes of $R_0/(e\mathcal{E}_z)$ and $\frac{A^2}{4M^2}$, one can see that close to the critical thickness $d_{QW} = d_c$ determining the transition from normal to topologically non-trivial insulator, the magnitude of the in-plane term is an order of magnitude larger than SOI term in z-direction, while for d=80 Å the magnitudes of both interactions are comparable.

Note also that the lowest order term related to bulk inversion asymmetry (BIA) of the QW has the form (in the basis of $\tilde{H}$)

$$\tilde{H}_{BIA} = \begin{pmatrix} & & & -\Delta \\ & & \Delta & \\ & \Delta & & \\ -\Delta & & & \end{pmatrix} \tag{7.30}$$

with $\Delta \approx 1.6$ meV [Kön+08]. After folding to the electron or hole band, this term does not give any spin-dependent term in the effective model. Higher order terms due to BIA exist [Dre55], but they will be cubic in k for the effective electron model.

7.3 Spin transport within an effective electron model

As described in detail in the previous section, the effective conduction band description of a HgTe quantum well (7.25) includes two different SOI terms: the Rashba spin-orbit (SO) coupling and the in-plane Pauli term. To understand the interplay of both SOIs, we will analyse here, the spin Hall conductance signal numerically using the tight binding version (i.e. a lattice model including only nearest neighbour hopping terms) of the Hamiltonian (7.25) within the Landauer-Büttiker formalism [Dat07]. Some background of this formalism has been given in Chapter 5.

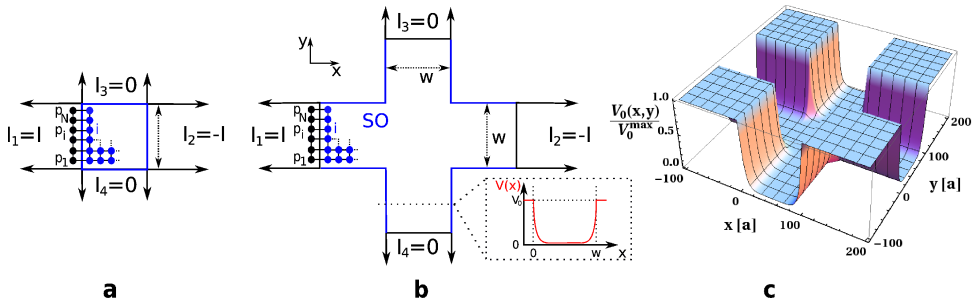


Figure 7.2: The two different samples we use for the numerics: (**a**) a quadratic Hall bar of width w and (**b**) a cross structure. In both cases the sample (blue) contains spin-orbit interaction, while the four semi-infinite leads (black) are SOI free. The numbering of the leads and the used boundary conditions on the currents are indicated in the figure, i.e. we drive a current I from lead 1 to 2 and do not allow for charge currents at the vertical leads 3 and 4. The discretization is shown for a part of each sample. Along the blue lines the confining potential $V_0(x, y)$ is applied, as indicated in the inset of (**b**), to give rise to the in-plane Pauli interaction. In case of the quadratic sample this potential corresponds to a tunnelling barrier between leads and sample. (**c**) shows the confining potential normalized by its maximal value (V_0^{max}) for the cross structure. Taken from [Rot+10]. © (2010) Creative Commons CC BY-NC-SA

7.3.1 Description of the model

The field $\mathcal{E}_z$ responsible for the Rashba term can be applied constant in space and varied in strength easily by an external top gate. By contrast $\mathcal{E}_\parallel$, generating the in plane Pauli term, usually originates from impurities or from the confinement due to sample boundaries. In our calculations we work in the quasi-ballistic regime, which is very well justified for HgTe/CdTe QWs with the typical mobilities 1 - 5 $\cdot$ 10^5 cm^2/Vs. Consequently the contribution of the impurities to the in-plane field is negligible and the confining potential is dominant, i.e. $|\mathcal{E}_{x,y}|$ decreases with the distance from the sample edges. The confining potential requires that an electric field at the boundary always points outside the sample, i.e. its magnitude changes sign at opposite edges.

We use two different setups. In both setups a finite size sample with spin-orbit interactions (see Figs. 7.2a,b) is attached to four semi-infinite leads of the same width w. For the first setup, a square sample, the in-plane field is introduced by tunnelling barriers between the leads and the conductor along the blue lines in the 7.2a. This simple setup has the advantage that the numerical results are easy to interpret, but is not a realistic description of actual experiments. The second setup is a symmetric cross structure, which resembles the experimental Hall bars and is shown in Fig. 7.2b. In this case again the blue lines indicate the sample border, where a confining potential is present (for a form of this potential see also Fig. 7.2c).

We construct the confining potential in two dimensions $V_0(x, y)$ from the one dimensional profile

$$V_{1D}(t) = e^{-t/l} + e^{-(w-t)/l} \tag{7.31}$$

where the coordinates in both the quadratic and the cross-shaped samples are chosen such that 0 respectively w mark the x- or y-coordinates at the edges of the central square. l is the

characteristic decay length of the potential. For the square-shaped sample, we define

$$V_0(x,y) = c_{sq}\left(V_{1D}(x) + V_{1D}(y)\right). \tag{7.32}$$

The maximal field is then $\mathcal{E}_x^0 = -\partial_x V_0(x,y)|_{x=0}$. We adjust the constant c_{sq} to choose some particular value $\mathcal{E}_x^0$. For the cross-shaped sample, we use the definition

$$V_0(x,y) = c_{cr} V_{1D}(\min(\max(x,0),w)) \cdot V_{1D}(\min(\max(y,0),w)). \tag{7.33}$$

Here the constant c_{cr} is adjusted to choose the desired maximal field $\mathcal{E}_x^0 = -\partial_x V_0(x,y)|_{x=y=0}$ and we assume that $\mathcal{E}_x^0 = \mathcal{E}_y^0$. Further, in the leads the potential is always set zero. For clarification, the confining potential corresponding to (7.33) is shown in Figure 7.2c. We find that the numerical results do not change qualitatively if the boundary field is defined differently, as long as the characteristic decay length l is unchanged. In both setups we assume that the leads do not include SO interactions, and therefore an analytical form of the eigenmodes [Sol+89] and a clear definition of the spin current is available [Han+04].
In the calculations, we set the boundary conditions on the currents $I_1 = -I_2 = I$ and $I_3 = I_4 = 0$, where $I_p = I_p^\uparrow + I_p^\downarrow$ is the total current at lead p. The spin-dependent current, I_p^σ, is calculated by use of the Landauer-Büttiker formula Eq. (5.54) [Büt+85; Han+04; Dat07]

$$I_p^\sigma = \frac{e}{h} \sum_{q \neq p} \sum_{\sigma'=\uparrow,\downarrow} \left[T_{qp}^{\sigma'\sigma} \mu_p - T_{pq}^{\sigma\sigma'} \mu_q \right], \tag{7.34}$$

which links the spin-resolved current to the chemical potential $\mu_p = eV_p$ via the transmission matrix elements $T_{pq}^{\sigma\sigma'}$. $T_{pq}^{\sigma\sigma'}$ describes the probability that an electron, entering the sample at lead p with spin σ, will leave the sample through lead q having spin z-projection σ'.

For a sample with non-zero SO coupling, applying an electric field between leads 1 and 2 will generate a transverse spin current $I_p^s = -\hbar/(2e)(I_p^\uparrow - I_p^\downarrow)$ at leads 3 and 4, which is the so called spin Hall effect [Han+04; Nik+05b; Han+05]. We define the spin Hall conductance as follows:

$$G_p^s = \frac{I_p^s}{V_1 - V_2} = \frac{\hbar}{2e} \frac{I_p^\uparrow - I_p^\downarrow}{V_2 - V_1} = \frac{e}{4\pi} \sum_{\sigma'=\uparrow,\downarrow} \left(T_{p1}^{\uparrow\sigma'} - T_{p1}^{\downarrow\sigma'} \right), \quad p = 3,4 \tag{7.35}$$

where $V_2 - V_1$ is the voltage difference between leads 2 and 1. Due to the symmetries of the Hamiltonian, only a few transmission matrix elements are independent, so that the last equality in (7.35) follows. The time reversal symmetry of (7.25) implies $T_{pq}^{\sigma\sigma'} = T_{qp}^{-\sigma'-\sigma}$, the fourfold rotational symmetry $\mathcal{C}_4$ of the setup about the z-axis implies e.g. $T_{23}^{\sigma\sigma'} = T_{14}^{\sigma\sigma'}$, and the mirror symmetry with respect to the yz-plane implies $T_{32}^{\sigma\sigma'} = T_{31}^{-\sigma-\sigma'}$.

The transmission matrix elements are computed numerically in a tight binding approach by using the Green's function method (Chapter 5 and [Dat07; Sol+89; Han+04; Han+05]), and the Fisher-Lee relation [Fis+81] connecting the Green's function with the transmission amplitudes. We discretize the sample as indicated in Fig. 7.2. By making use of the fermionic field operators $c_{\alpha,\sigma}^\dagger$ ($c_{\alpha,\sigma}$), which create (annihilate) a spin σ electron at lattice site α, the Rashba and in-plane Pauli interactions take the following form in second quantization:

$$\mathcal{H}_{\text{Rashba}} = \frac{R_0}{2a} \sum_\alpha \left[\mathrm{i} c_{\alpha,\uparrow}^\dagger c_{\alpha+a_y,\downarrow} + \mathrm{i} c_{\alpha,\downarrow}^\dagger c_{\alpha+a_y,\uparrow} - c_{\alpha\uparrow}^\dagger c_{\alpha+a_x\downarrow} + c_{\alpha\downarrow}^\dagger c_{\alpha+a_x\uparrow} + \text{h.c.} \right] \tag{7.36}$$

$$\mathcal{H}_{\text{Pauli}} = \mathrm{i} \frac{A^2 e}{a 8 M^2} \sum_{\alpha,\sigma} \left[\mathcal{E}_x^{\alpha+a_y/2} c_{\alpha+a_y,\sigma}^\dagger c_{\alpha,\sigma} - \mathcal{E}_y^{\alpha+a_x/2} c_{\alpha+a_x,\sigma}^\dagger c_{\alpha,\sigma} - \text{h.c.} \right] \kappa_\sigma \tag{7.37}$$

where $R_0/(e\mathcal{E}_z)$, A and M are the material parameters defined in Section 7.1.1. Here a denotes the lattice constant, and $a_{x,y}$ stand for the lattice unit vectors connecting nearest neighbours. To obtain the parameter R_0 we must assume a specific value of the perpendicular field $\mathcal{E}_z$. $\mathcal{E}_x^\alpha$ and $\mathcal{E}_y^\alpha$ designate, for a given site α, the in-plane electric field components. $\kappa_\sigma = \pm 1$ for spin up and down, respectively, and the symbol h.c. denotes the Hermitian conjugate. The Rashba spin-orbit interaction does not conserve the z-component of the spin and thus leads to spin precession [Nik+05a]. By contrast, the in-plane Pauli term (7.37) conserves the z-component of the spin, causing a shift in energy for two spin directions. This energy shift however must not be mistaken as Zeeman effect, because (7.37) does not break time reversal symmetry.

For numerical calculations we consider the Hamiltonian

$$\mathcal{H} = \hat{T}_k + \mathcal{H}_{\mathrm{SO}} + \mathcal{H}_{\mathrm{Dis}} + V_0(x,y) + \mathcal{H}_{\mathrm{Dar}}. \tag{7.38}$$

Here $\hat{T}_k = \big(-D - B + A^2/(2M)\big)\cdot k^2 = \hbar^2 k^2/(2m^*)$ describes the kinetic part of the conduction band Hamiltonian (7.25). In second quantization, $\hat{T}_k$ is described by spin-conserving nearest neighbour hopping [Sol+89; Dat07; Han+04]. $\mathcal{H}_{\mathrm{Dis}} = \mathrm{diag}(\varepsilon_i)$ specifies the disorder of the sample, where the diagonal on-site energies ε_i are uniformly distributed between $[-W/2, W/2]$ [She+05]. The disorder strength $W = \hbar e/(m^*\mu)$ is calculated from the mobility μ. The confining potential is taken into account via $V_0(x,y)$. The spin-orbit coupling $\mathcal{H}_{\mathrm{SO}}$ is described by the Rashba term (7.36), the in-plane Pauli term (7.37) or a linear superposition of both terms. Hence $\mathcal{H}$ mirrors the conduction band Hamiltonian (7.25), where we omit the k-dependence of the $\mathcal{R}$ parameter and negligible terms which include the combined effect of the in-plane and out of plane electric fields. The spin-independent Darwin term $\mathcal{H}_{\mathrm{Dar}} \propto \nabla \cdot \mathcal{E}_\parallel$ breaks the particle-hole symmetry of the tight binding Hamiltonian, just like any space dependent in-plane potential would do. Here particle-hole symmetry means the relation $G_3^s(E_f) = -G_3^s(-E_f)$ if the energy zero point is chosen in the middle of the tight binding band. It originates from the cosine dispersion relation of a free electron on a lattice. $\mathcal{H}_{\mathrm{SO}}$ does not break this symmetry. The Darwin term does not qualitatively change the spin conductance signal and will be considered after the spin-orbit terms are analysed.

We use realistic parameters for the calculations, which are shown in Table 7.2. Here we assume a thickness of the quantum well in z-direction of 7 nm, corresponding to the inverted regime. Although this has no impact in our one band approach, it guarantees a large coupling strength $R_0/(e\mathcal{E}_z)$. Note that our parameters correspond to a regime where the gap of the band structure is found at $k = 0$, i.e. it is a direct gap. Therefore, the spin-orbit related parameter A gives a strong correction to the effective mass of the models $\hat{H}_e$ and $\hat{H}_h$, which changes the sign. The carrier density is set to $n = 2 \cdot 10^{10}$ cm^{-2} while the effective mass $m^* = 0.00712\, m_0$, where m_0 is the bare electron mass. We want to point out that we have changed the originally negative sign of the effective mass to achieve a positive curvature of the band dispersion, which is justified in an one band model. In our calculations, we focus on the regime where the Fermi energy is deep enough in the electron band (that is described by $\hat{H}_e$) so that the spin edge states have already merged into the bulk [Kön+08; Zho+08] and the spin Hall conductance which is the property of the states close to the Fermi level will be dominated by the bulk response.

Assuming a quadratic dispersion around $\mathbf{k} = 0$, we determine the Fermi energy to be $E_f = 6.73$ meV, by the relation to the density $n = \frac{m^*}{\pi\hbar^2} E_f$ of the 2DEG.[4] Finally the assumed mobility $\mu = 25 \cdot 10^4$ cm^2/(Vs) leads to $W = 0.65$ meV. For such a small disorder strength averaging over

[4] If we want to compare spin Hall signals of this chapter with results of Chapter 9, we should express the Fermi energy in terms of the hopping energy of the numerical model, $t_0 = \frac{\hbar^2}{2m^* a} = 0.214$ eV , with the lattice constant a.

10 different disorder configurations is sufficient. We note that the mean spin Hall conductance deviates from that of a clean sample only by about $\lesssim 1\%$. The parameters are chosen carefully to the restricted range of possible energies where this model is valid, i.e. $E_{\mathrm{r}} < E_f < E_{\mathrm{gap}}$, where $E_{\mathrm{gap}} = 2|M|$ is the energy gap and E_{r} is the energy splitting of the band due to Rashba interaction.

We will first focus on the square sample because the scattering barrier in this setup allows to study the competition between the in-plane and out-of plane electric fields. The influence of in-plane electric fields is much weaker in the cross structures. Further, the first minimum in the spin Hall conductance generated by the Rashba contribution is shifted away from the spin-precession length to smaller fields due to quantum interference effects in the vertical stubs (see e.g. [Sol+89]) and therefore results are less transparent to interpret.

7.3.2 Numerical results for the quadratic sample

We choose a quadratic sample of width $w = 1000$ nm, which is discretized by 200×200 lattice points, so that the Fermi wavelength is about 36 times the lattice constant. The characteristic length scale of the electric field is assumed to be $l = 10$ nm. The computed spin Hall conductance is presented in Figure 7.3. First we focus on pure SOI and consequently omit the Darwin term and the potential in Figures 7.3a and b. Figure 7.3a shows $G_3^s(\mathcal{E}_{\parallel}^0)$ for different top gate fields and Figure 7.3b presents $G_3^s(\mathcal{E}_z)$ for different in-plane fields.

Rashba coupling: The spin Hall conductance signal induced by the Rashba coupling alone, $\mathcal{E}_{\parallel} = 0$, is shown in Figure 7.3b by blue circles. For a small interaction strength R_0, the spin Hall conductance rises quadratically, saturates and finally starts to precess. The behaviour of the spin Hall conductance originating from the Rashba model can be understood by the spin force operator

$$\hat{\mathbf{F}}_H = \frac{-m^*}{\hbar^2}[[\hat{\mathbf{r}}_H, \mathcal{H}], \mathcal{H}] = \frac{2m^* R_0^2}{\hbar^3} (\hat{\mathbf{p}}_H \times \mathbf{z}) \otimes \hat{\sigma}_H^z \tag{7.39}$$

where $\hat{\mathbf{r}}_H$, $\hat{\mathbf{p}}_H$ and $\hat{\sigma}_H^z$ are the position, momentum and spin-operators in the Heisenberg representation (see Nikolić et al. [Nik+05a]) and $\mathbf{z}$ is a unit vector. $\hat{\mathbf{F}}_H$ does not depend on $\mathcal{E}_{x,y}$. In this simple picture the force acting on electrons due to SO coupling is quadratic in R_0, explaining the behaviour of G_3^s as a function of the out-of plane electric field for low $\mathcal{E}_z$. The force described in (7.39) deflects the spin-$\uparrow$ and the spin-$\downarrow$ electrons in opposite transverse directions leading to the spin Hall effect. However, the Hamiltonian does not conserve the z-component of the spin, leading to a rotation of the spin direction and as a consequence to oscillations in G_3^s as a function of R_0. The first maximum of the spin Hall conductance is reached, when the spin has travelled a distance equivalent to the spin precession length [Nik+05b] $L_{SO} = \pi \hat{T}_k/(k^2 R_0)$, over which the expectation value $\langle \sigma_z \rangle$ rotates by π. The electric field $\mathcal{E}_z^{SO} = 1.08$ mV/nm corresponding to the expected maximum is indicated as a red line in Figure 7.3b and is in good agreement with the maximum of the absolute value of G_3^s.

In-plane Pauli interaction: The spin Hall conductance shows a linear behaviour as a function of the in-plane electric field (see blue circles in Figure 7.3a, where only the $\mathcal{E}_{x,y}$ components are non-zero). This linear dependence on field strength can be explained within the

Here we have $E_f = 0.03t_0$, both for the square and cross samples. Since in this chapter, we analyse only the effective model $\hat{H}_e$ of the electron band, and take band parameters of the inverted regime, we can compare the spin Hall conductance for $\mathcal{E}_z = 0$ in Figs. 7.3 with the red solid line in Fig. 9.9a, E_f slightly below the bulk gap. This shows that the in-plane potential of this chapter should be quite strong, in order to model the effect of the confinement on the spin Hall conductance of the four-band model.

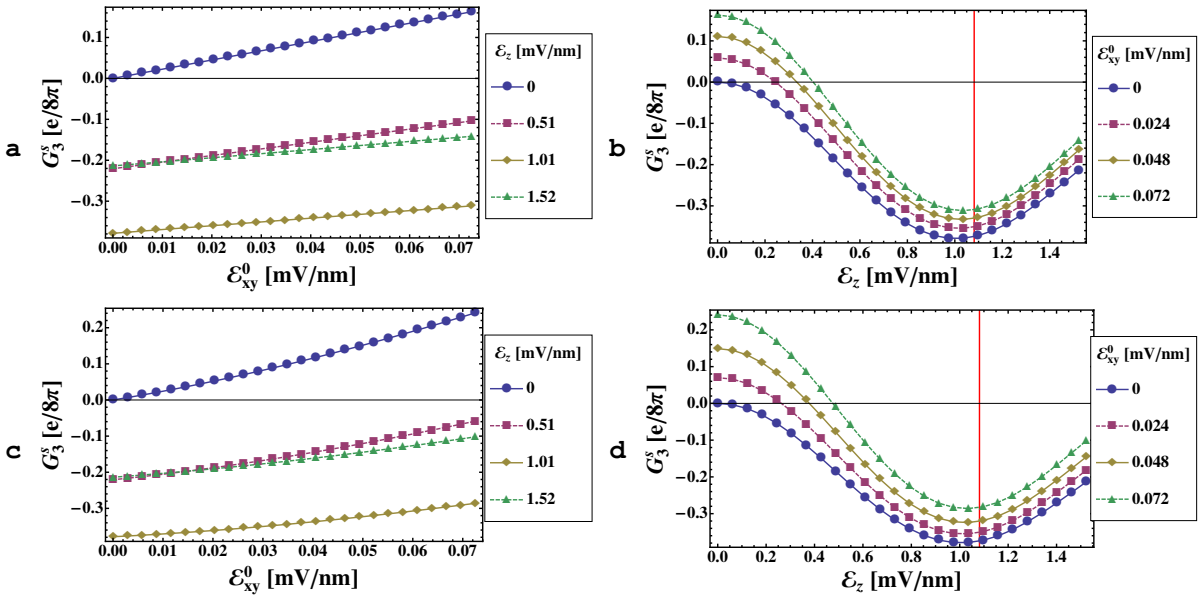


Figure 7.3: The spin Hall conductance due to the presence of a superposition of Rashba and in-plane Pauli terms in a quadratic sample with tunnelling barriers at the boundaries. Due to the presence of the second interaction the starting value of spin Hall conductance can be non-zero. (**a**) When $\mathcal{E}_z$ is non-zero, the magnitude of G_3^s as function of $\mathcal{E}_{x,y}^0$ is reduced. (**b**) The spin Hall conductance as a function of $\mathcal{E}_z$. For higher values $\mathcal{E}_{x,y}^0$ the precession amplitude of G_3^s at L_{SO} (indicated by a red line) is increased and the minimum is slightly shifted to higher interaction energies. (**c**) and (**d**) show the same dependencies of G_3^s as above, but here the Darwin term and the confining potential were additionally taken into account. Taken from [Rot+10]. © (2010) Creative Commons CC BY-NC-SA

semiclassical approach, where we have adopted the wave packet dynamics by Sundaram and Niu [Sun+99] (compare Chapter 4.1) to obtain equations of motion for the in-plane Pauli term:

$$\dot{\mathbf{r}}_c = \frac{1}{\hbar}\frac{\partial \epsilon}{\partial \mathbf{k}_c} - \dot{\mathbf{k}}_c \times \mathbf{\Omega}_\sigma \tag{7.40}$$

$$\hbar\dot{\mathbf{k}}_c = -e\mathcal{E}_\parallel, \tag{7.41}$$

with the magnetic field set to zero. Here the index c denotes the center coordinate of the wave packet in position and $\mathbf{k}$ space. The Berry curvature is defined as

$$\left(\mathbf{\Omega}_\sigma^\pm\right)_\alpha := -\varepsilon_{\alpha\beta\gamma}\mathrm{Im}\left\langle \frac{\partial u_\sigma^\pm(\mathbf{k})}{\partial k_\beta} \middle| \frac{\partial u_\sigma^\pm(\mathbf{k})}{\partial k_\gamma} \right\rangle, \tag{7.42}$$

where the symbol $\pm$ corresponds to two eigenvalues $E_\pm = -Dk^2 \pm \sqrt{A^2k^2 + \mathcal{M}^2(k)}$ of the upper spin block of the Hamiltonian $\tilde{H}_0$ in (7.12) with $u_\sigma^\pm(\mathbf{k})$ being corresponding eigenstates for spin σ. Since (7.12) does not include the Rashba SO interaction, the Berry curvature is diagonal in the spin index $\sigma = \uparrow, \downarrow$, and non-Abelian corrections are not present. (7.41) simply describes the change of the lattice momentum due to the electric field. (7.40) describes the time evolution of the position operator due to the band dispersion and the anomalous velocity term $\dot{\mathbf{k}}_c \times \mathbf{\Omega}_\sigma$ with

the spin-dependent non-zero z-components $(\mathbf{\Omega}^{\pm}_{\uparrow})_z = -(\mathbf{\Omega}^{\pm}_{\downarrow})_z \overset{\mathbf{k}=0}{=} \pm\frac{A^2}{2M^2}$, compare Eq. (4.49) for the anomalous velocity in the Dirac model. Compared to the prefactor $\mathcal{G} = \frac{A^2}{4M^2}$ of the SO term in the Pauli equation (7.25), the prefactor $\frac{A^2}{2M^2}$ of the anomalous velocity is two times larger. In Section 4.2 we showed that the difference is connected with a non-commuting position operator, and we interpreted the correction as fingerprint of the relativistic Thomas-Wigner precession. The spin-dependent anomalous velocity term shifts the position of the wave packet with different spins in two opposite transverse directions leading to the spin Hall effect. Inserting (7.41) into (7.40) yields the dependence of the anomalous velocity term linear in $\mathcal{E}_{\parallel}$. Note, that the energy range (up to $\approx 0.35E_f$) plotted in Figures 7.3a and b is the same for both interaction terms. Since the coupling parameters can be quite different (in our calculations $21 \cdot R_0/(e\mathcal{E}_z) \approx A^2/(4M^2)$), the magnitudes of in-plane and out-of plane electric fields are adjusted so the interaction energies are the same, allowing us to analyse the interplay of both interactions.

Interplay of both interactions: A linear superposition of (7.36), (7.37) and $\hat{T}_k$ leads to the spin signal which is shown in Figures 7.3a and b, when all three field components are non-zero. The finite value of the spin Hall conductance in Figure 7.3a for $\mathcal{E}^0_{x,y} = 0$ is due to the Rashba coupling. It can be observed, that the linear behaviour of the in-plane Pauli term with the electric field is not changed, when Rashba spin-orbit coupling is present. However, the slope of the spin Hall conductance curves decreases with $\mathcal{E}_z \neq 0$, which means that the in-plane Pauli contribution to the spin signal is suppressed by the Rashba interaction. The z-component of the spin is not conserved for finite $\mathcal{E}_z$, as can be seen from Equation (7.36). The resulting spin precession implies that generation of a spin current by the anomalous velocity becomes less effective. The smallest slope and therefore smallest in-plane Pauli contribution in Fig. 7.3a is found for an electric field $\mathcal{E}_z$ corresponding to matching sample length and precession length.

Figure 7.3b shows G^s_3 as a function of $\mathcal{E}_z$ for different in-plane electric fields. One can see the typical precession pattern of $G^s_3(\mathcal{E}_z)$ also for $\mathcal{E}_{\parallel} \neq 0$. Moreover the precession amplitude of the spin Hall conductance of the Rashba type is enhanced in the presence of the in-plane Pauli term. The origin of this increase can be traced back to the **k**-dependent energy splitting of the spin subbands due to the in-plane Pauli interaction. In order to lower its energy the electron now prefers to stay in either spin up or down states. The precession of the spin is thus slightly suppressed, as can be seen in a small shift of the minima to higher electric fields. As discussed above, the spin force operator (7.39) can act more efficiently on electrons with a preferred spin z-projection, which leads to the relatively higher magnitude of the spin Hall conductance caused by Rashba coupling.

Figs. 7.3c and 7.3d also show $G^s_3(\mathcal{E}_{\parallel})$ and $G^s_3(\mathcal{E}_z)$ respectively but now with included potential $V_0(x, y)$ and Darwin $A^2/(8M^2)\nabla \cdot \mathcal{E}_{\parallel}$ terms. In the tight binding approach, both terms renormalize the diagonal on-site energy. They cannot generally be considered to be small, as they scale, like the in-plane Pauli term, with the magnitude and the shape of the confining electric field. The relative magnitudes of the in-plane Pauli, Darwin and potential terms depend on the choice of the functional dependence $V_0(x, y)$. The most important scale is the characteristic length scale l, over which the corresponding field drops to $\mathcal{E}^0/e$. We have performed numerical calculations with different values $\mathcal{E}^0_{x,y}$ and l and found, that the main features of the results discussed in this paper stay the same. We found that, the Darwin term is the main contribution which renormalizes the spin Hall conductances in Figs. 7.3c and 7.3d in respect to $G^s_3(\mathcal{E}_{\parallel})$ and $G^s_3(\mathcal{E}_z)$ in Figs. 7.3a and 7.3b.

The divergence of $\mathcal{E}_{\parallel}$ appears as an additional term in the semiclassical equation (7.41) and therefore the spin-independent Darwin term can contribute to the anomalous velocity and renor-

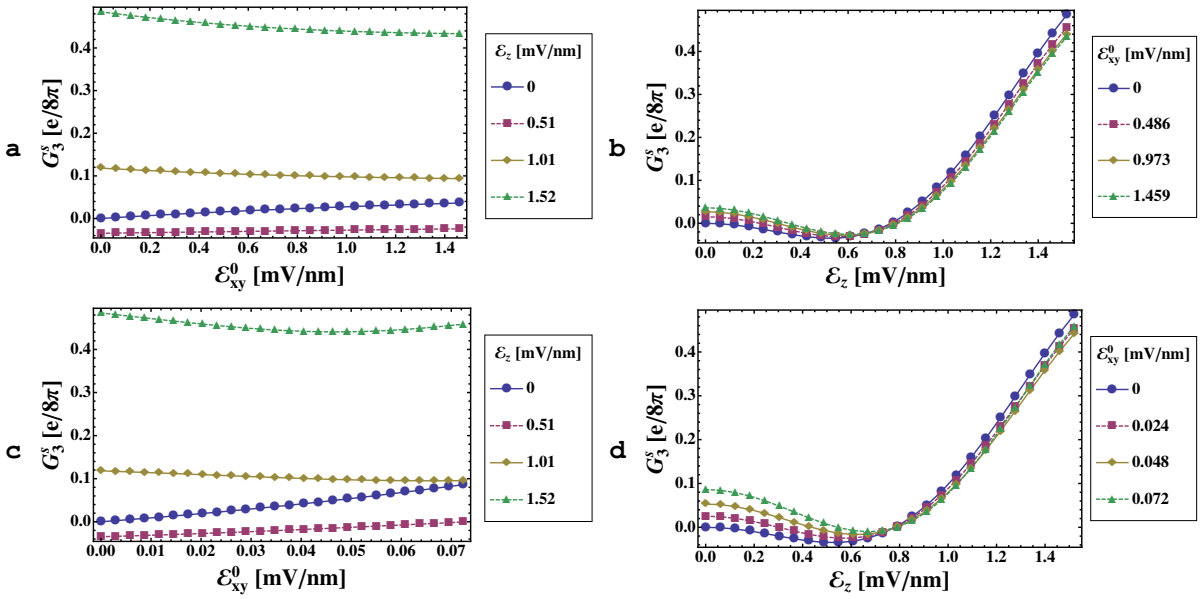


Figure 7.4: The spin Hall conductance for a cross structure, with the Darwin and the potential terms included. In (**a**) and (**b**) we have used the same range of influence of the electric field as for the quadratic sample, i.e. $l = 10$ nm. Although the in-plane field was enhanced to very high values, the influence of in-plane electric field on $G_3^s(\mathcal{E}_z)$ is very weak. In (**c**) and (**d**) we have increased l to 40 nm to get a higher in-plane Pauli signal. In this case G_3^s behaves similarly as for the quadratic sample. Taken from [Rot+10]. © (2010) Creative Commons CC BY-NC-SA

malise the spin Hall conductance term. This can be seen in Fig. 7.3c as a non linear behaviour of $G_3^s(\mathcal{E}_\parallel)$. However, in the range of in-plane electric fields shown in Figs. 7.3c and 7.3d, the qualitative behaviour of the spin Hall conductance is the same as in the absence of the Darwin term (see Fig. 7.3a,b). Increasing the in-plane electric field to the same magnitude as the electric field perpendicular to the 2DEG bears two difficulties. First of all the interaction energy of the in-plane Pauli term will exceed the Fermi energy which marks the limit of validity of our effective electron model. Secondly, increasing $\mathcal{E}_\parallel$ comes along with a raising tunnelling barrier in the quadratic sample. We can omit these difficulties by choosing the sample in a shape of a cross (see next subsection).

7.3.3 Numerical results for the cross sample

The cross sample is constructed of 5 square parts: 4 stubs and the central square (see Fig. 7.2b). Each part has the width $w = 500$ nm and is discretized by 100×100 lattice sites. The corresponding spin Hall conductance originating from in-plane Pauli and Rashba terms in the presence of the Darwin term and the confining potential is shown in Figs. 7.4a - d.

Figures 7.4a,b show $G_3^s(\mathcal{E}_\parallel)$ and $G_3^s(\mathcal{E}_z)$ for different values of fields in z and in-plane directions correspondingly and for the characteristic range of the electric field $l = 10$ nm. These figures should be compared with Figs. 7.3c and 7.3d correspondingly. The overall behaviour of spin Hall conductance in Figs. 7.4a and b is similar to the quadratic sample. However, although we used values of in-plane electric field around twenty times larger than in Figures 7.3d the influence

of $\mathcal{E}_\parallel$ on $G_3^s(\mathcal{E}_z)$ is much weaker than for the square structure (compare 7.4b with 7.3d). The probability of scattering from vertical and horizontal walls in the cross structure is much smaller than in the case of the square structure where the electron directly hits the wall. Therefore we find that the influence of in-plane electric field will be much weaker in the experimentally relevant cross samples.

In order to obtain a contribution to G_3^s due to the confining potential comparable to that for the quadratic sample, we have increased l from 10 nm to 40 nm, in Figs. 7.4c,d. This leads to a larger range of lattice sites, which can contribute efficiently to the spin dependent hopping. In the case of the larger l the results resemble those discussed in the last section for square samples. In Figure 7.4d the amplitude of the Rashba contribution to the spin Hall effect counted to the first minimum is enhanced by the confinement potential. By contrast the signal due to the in-plane Pauli term decreases until $\mathcal{E}_z$ corresponds to the first minimum as can be seen in Fig. 7.4c. Similar to our observation in Fig. 7.3c, the slope $\partial G_3^s/\partial \mathcal{E}_{xy}^0$ decreases with increasing top gate field, as long as the spin precession angle due to $\mathcal{E}_z$ is smaller than π.

At the end of this section let us emphasize, that for the experimentally relevant case both interactions are present, but only $\mathcal{E}_z$ can be easily varied, e.g. by a top gate. Therefore, in the experimentally relevant case, the presence of both the in-plane Pauli and the Darwin interactions could lead to an increase of the amplitude of spin signal.

7.4 Conclusions

We have derived an extension to the BHZ Hamiltonian for the typical 2D topological insulator (HgTe QWs) in the presence of the inversion breaking potential in z-direction, and an in-plane potential. For the derivation, we used two independent methods: $\mathbf{k}\cdot\mathbf{p}$ perturbation theory and symmetry arguments based on Clifford algebra. We found that to the third order in the perturbation theory, only the inversion breaking potential in z-direction generates new off-diagonal in spin space terms. These terms lead to the Rashba spin-orbit interaction when the Foldy-Wouthuysen transformation to the effective electron model is performed. On the other hand the diagonal-in-spin space part of the Hamiltonian in the presence of the in-plane potential generates an additional term to the one band model which is also linear in momentum and spin, but conserves the z-component of the spin. By analogy with the equation for a relativistic electron in vacuum we call this term in-plane Pauli term. The presence of both terms in the conduction band Hamiltonian leads to an interesting behaviour of the spin Hall conductance. In particular, the in-plane Pauli contribution to the spin Hall conductance is suppressed in the presence of the spin precession inducing terms. By contrast, the spin Hall conductance from the Rashba term preserves the oscillation pattern in the presence of the in-plane Pauli term and its magnitude can be enhanced due to partial pinning of the z-component of the spin. This latter situation is experimentally relevant since the inversion breaking potential in z-direction can be easily tuned by a top gate in experiments. Therefore we expect that in experiments on asymmetrically doped HgTe/CdTe QWs [Brü+10] in the metallic regime (the Fermi level in the conduction or valence band), the behaviour of spin transport and especially the spin Hall conductance will be dominated by the Rashba spin-orbit interaction. Note, that in our derivation we omit the BIA terms since they are already studied in Ref. [Kön+07], and since at lowest order, they do not add new spin physics to the effective electron or hole band models.

Let us also emphasize that our effective four band Hamiltonian in the presence of inversion breaking potential is not limited only to the HgTe/CdTe QWs and can be easily generalized to

other topological insulators like type II InAs/GaSb/AlSb quantum wells [Liu+08] with correctly adjusted strengths of the Rashba spin-orbit interactions.

Chapter 8

All-electric qubit control in heavy hole quantum dots via non-Abelian geometric phases

In this chapter, we show an application of the adiabatic theorem and the associated non-Abelian holonomy to the universal control of a heavy hole (HH) spin qubit. The adiabatic theorem of quantum mechanics and the associated geometric or Berry phase, and also its generalization to degenerate subspaces, which can lead to a non-Abelian holonomy, have been discussed in Chapter 4. Further, in Chapter 6 we have seen that a special case of the holonomy, given by an integral over the Berry curvature, is related to the first Chern number, which is a topological invariant. In this sense, non-trivial holonomy is linked to non-trivial topology. In this chapter we will be interested in the holonomy, but will not further elaborate on topological aspects. While all the chapters of this thesis are concerned with manipulation of spins or spin currents in general, and thus contribute to the area of spintronics, this chapter is different in the sense that control of a single spin is a specific requirement for spin based quantum computation. The idea of building a quantum computer [Fey82] has attracted a lot of interest due to the vast increase in computation power which could become possible for certain algorithms, the most interesting of them being applications in cryptography, and the simulation of quantum physics itself [Nie+10]. One promising technological path for implementing a quantum computer is based on spin qubits in quantum dots. Therefore, the device shown here may serve as computation gate in the proposed scheme of universal quantum computation by Loss and DiVincenzo [Los+98].

This chapter is based on our publication [Bud+12a], where we have demonstrated universal control of a spin qubit in heavy hole quantum dots in the absence of magnetic fields. A time dependent electric quadrupole field is used to perform any desired single qubit operation by virtue of the holonomy. Not relying on magnetic fields is a clear advantage with respect to technological integration of quantum gates. Another important advantage is that the presence of time reversal symmetry (TRS) is known to forbid several dephasing mechanisms, for example, in HH spin qubits due to the interplay of electron phonon coupling and Rashba spin orbit coupling [Bul+05]. Therefore, coherent spin control by all-electric means is one of the major goals of spintronics. In the original work [Los+98], the proposed scheme for universal quantum computing based on spin qubits in quantum dots (QDs) relied, on the one hand, on all-electric two qubit operations but, on the other hand, on single qubit operations based on magnetic fields or ferromagnetic auxiliary devices that both break TRS. A few years later, electric-dipole-induced

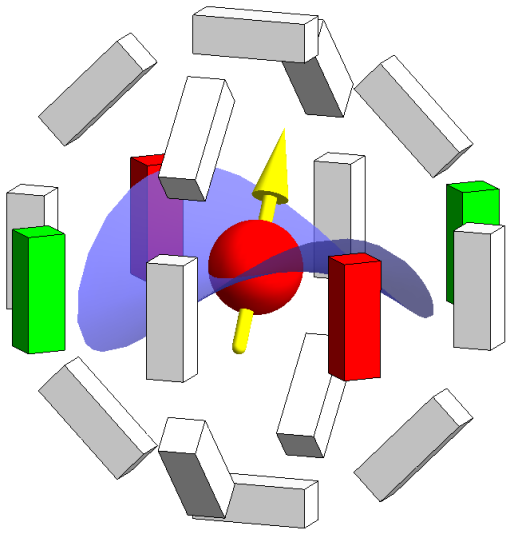


Figure 8.1: Schematic of single particle (red ball) with a HH (pseudo)spin (yellow arrow) in a $J = \frac{3}{2}$ valence band QD. The three-dimensional QD is surrounded by 18 gates that allow to generate an electrostatic potential with quadrupole symmetry in any direction in space. The red and green colors on the gates should visualize applied electrostatic potentials with opposite sign that give rise to the quadrupole field shown in light blue. Taken from [Bud+12a]. © (2012) by the American Physical Society

spin resonance (EDSR) has been proposed [Gol+06] and experimentally realized [Now+07] as a way to process spins electrically in presence of a static magnetic field which is still breaking TRS. Rather recently, it has been theoretically shown that in spin qubits based on carbon nanotube QDs it is indeed possible to accomplish all-electric single qubit operations using EDSR [Bul+08; Kli+11]. This is true because the specific spin orbit interaction in carbon nanotubes provides a way to split spin up and spin down states in the absence of magnetic fields. However, spin qubits based on carbon nanotubes face other problems and it is fair to say that all host materials for spin qubits have advantages and disadvantages.

In this work, we show how universal single qubit operations can be performed by all-electric means in the framework of holonomic quantum computing [Zan+99] in HH QD systems. The adiabatic evolution in presence of a time dependent electric quadrupole field is employed to control the HH qubit (see Fig. 8.1 for a schematic). For our purposes, HH spin qubits (composed of $J = \frac{3}{2}$ states) are the simplest two level system that can be manipulated in the desired way. However, HH spin qubits are, of course, a very active research area by itself beyond holonomic quantum computing. Two reasons why HH QDs are promising and interesting candidates for spin qubits are, for instance, the advanced level of optical control [Ger+08; Ebl+09; Bru+09; deGre+11] and the predicted long coherence times [Fis+10].

The topological properties of TRS preserving half integer spin systems have been analysed in a series of seminal papers by Avron and co-workers [Avr+88; Avr+89]. The case $J = \frac{3}{2}$ is of particular interest both from a theoretical and from a more applied point of view. From the theoretical side, all TRS preserving gapped Hamiltonians are unitarily related due to an SO(5) symmetry [Avr+88; Avr+89] giving rise to an SO(5) Clifford algebra [Avr+89; Dem+99] which allows for a simple analytical calculation of the adiabatic time evolution and with that

the geometric phase. From the experimental side, the $J = \frac{3}{2}$ system is naturally realized in the p-like valence band of many semiconductors where spin orbit coupling isolates the $J = \frac{3}{2}$ states from the so called split-off band. Interestingly, the fingerprints of $SU(2)$ non-Abelian geometric phases [Kat50; Wil+84] could be also identified in several transport properties of this class of semiconducting materials [Aro+98; Mur+04].

The pioneering idea of using non-Abelian holonomy to perform quantum computing tasks in the $J = \frac{3}{2}$ system is due to Bernevig and Zhang [Ber+05] who proposed the electric Stark effect to process valence band impurities in III-V semiconductors. The idea works for the light hole (LH) subspace of the $J = \frac{3}{2}$ Hilbert space. However, the resulting holonomy is Abelian on the HH subspace [Zee88] so that the electric Stark effect cannot be used to process HH qubits. In contrast, the electric quadrupole fields employed in our proposal allow for full adiabatic control over the entire $J = \frac{3}{2}$ Hilbert space. As we will see, the effective quadrupole Hamiltonian is mathematically a generalization of the Stark Hamiltonian, so our method of calculation could also be used to find the holonomy that is generated by control over the electric field vector in a system showing the (quadratic) Stark effect.

Recently, holonomic quantum computing due to tunable spin orbit coupling with electron spins in spatially transported quantum dots has been suggested [San+08; Gol+10] but this is very demanding from an experimental point of view. Our idea is conceptually much simpler. We derive below the time dependent electric quadrupole field that realizes any desired single qubit operation

$$\mathcal{U}(\hat{n}, \varphi) = \exp\left(i\varphi \frac{\hat{n}\sigma}{2}\right) \tag{8.1}$$

on the HH spin qubit. Here, $\hat{n}$ is a unit vector representing the rotation axis, φ is the angle of the rotation, and σ denotes the vector of Pauli matrices acting on the qubit space. Furthermore, we give an estimate of the adiabatic time scale which determines the maximum operating frequency of single qubit gates showing that the physics we describe is experimentally feasible. Finally, we discuss the influence of several imperfections, which might be present in an experimental setup, on the working precision of our proposal.

The non-Abelian geometric phase, occurring in a degenerate subspace after a cyclic evolution, is readily expressed once the time-dependent projection $P(t)$ onto this degenerate subspace is known. The generator of the adiabatic evolution (compare Chapter 4) then reads [Kat50]

$$\mathcal{A}\left(\frac{d}{dt}\right) = -\left[\frac{dP(t)}{dt}, P(t)\right]. \tag{8.2}$$

On the basis of this generator, the non-Abelian geometric phase [Wil+84] associated with a loop γ in parameter space is given by the holonomy

$$U_\gamma = \mathcal{T} \mathrm{e}^{\int_\gamma \mathcal{A}}, \tag{8.3}$$

where $\mathcal{T}$ denotes time-ordering. For the Hilbert space of a $J = \frac{3}{2}$ particle in presence of TRS this holonomy is readily calculated analytically as we explicitly demonstrate below.

In Chapter 3 about the theory of invariants, we have already seen that the Hamiltonian of a spin-$\frac{3}{2}$ particle coupled to an electric quadrupole field can be written as [Avr+88]

$$H(\mathcal{Q}) = J_i \mathcal{Q}^{ij} J_j, \tag{8.4}$$

where J_i are components of the angular momentum operator and $\mathcal{Q}$ is the quadrupole tensor of the applied field. We put $\hbar = 1$ in the following. $\mathcal{Q}$ is a real, symmetric, traceless matrix. The space of such matrices is five dimensional. An orthonormal basis of this space is given by the matrices $\{Q_\mu\}_\mu$, $\mu = 0, \ldots, 4$, which satisfy the orthogonality relation $\frac{3}{2}\mathrm{Tr}\{Q_\mu, Q_\nu\} = \delta_{\mu\nu}$. We have given an explicit basis in Eq. (3.25). A general quadrupole tensor is then of the form $\mathcal{Q} = x^\mu Q_\mu$ and the associated Hamiltonian reads

$$H(\mathcal{Q}) = H(x^\mu Q_\mu) = x^\mu J_i Q_\mu^{ij} J_j \equiv x^\mu \Gamma_\mu, \tag{8.5}$$

where the basis Hamiltonians $\Gamma_\mu \equiv J_i Q_\mu^{ij} J_j$ obey the SO(5) Clifford algebra[1] [Avr+89; Dem+99]

$$\{\Gamma_\mu, \Gamma_\nu\} = 2\delta_{\mu\nu}. \tag{8.6}$$

Although this relation looks similar to the orthogonality relation of the Q_μ, there is no trivial connection between the two. From the Clifford algebra property (8.6), it is clear that the eigenenergies of $H(\mathcal{Q})$ are $\pm|\mathbf{x}|$, each twofold (Kramers) degenerate.

Let us shortly discuss how the effective quadrupole Hamiltonian (8.4) includes the effective Hamiltonian for the Stark effect in an acceptor centre or quantum dot in a T_d or O_h crystal. That is, we are interested in the splitting of the lowest bound electronic states at the impurity or quantum dot, due to an applied electrical field. In Chapter 3 we have shown how to construct the relevant Hamiltonian H_{E^2} of the quadratic Stark effect, see Eq. (3.28). We see that $H_{E^2} = \mathbf{J}^T \mathcal{Q}(\mathbf{E})\mathbf{J}$ with

$$\mathcal{Q}(\mathbf{E}) = \beta\left(Q_0 \frac{1}{\sqrt{3}}(2E_z^2 - E_x^2 - E_y^2) + Q_3(E_x^2 - E_y^2)\right) + \delta\left(Q_2 E_y E_z + Q_1 E_z E_x + Q_4 E_x E_y\right), \tag{8.7}$$

where the parameters β and δ are measures of polarisability. This shows that even in the anisotropic case, when $\beta \neq \delta$, for every path $\mathbf{E}(t)$ parametrizing the adiabatic change of the electric field vector with time, we can find a corresponding path $\mathcal{Q}(\mathbf{E}(t))$ in the quadrupole parameter space, which leads to identical holonomy acting on the spin. Therefore, the calculation of the holonomy for quadrupole fields, as shown in this section, can be seen as generalization of the calculation of holonomy for the quadratic Stark effect. The reverse is not true, since there are holonomy operations that can be obtained by changing the quadrupole field, but not by changing the electric field. We find[2] that universal holonomy operations, i.e. operations generating the full group SU(2), are possible with the quadratic Stark effect on the LH sector, but not on the HH sector, since there, the holonomy is Abelian. Also, the linear Stark Hamiltonian $H_{\mathbf{E}}$, (3.27), is not suitable for universal control of a spin, since the associated holonomy generated by adiabatic change of $\mathbf{E}$ is an Abelian subgroup of SU(2). Note also, that the mapping of the quadrupole Hamiltonian to a Stark Hamiltonian, as employed here, is only used to find an identification at the level of an effective Hamiltonian. Clearly, the physical situation is very different, since a quadrupole potential does not correspond to a homogeneous electrical field, although one could use the setup of quadrupole gates of Fig. 8.1 to generate electrical dipole fields.

As far as the geometric phase associated with a cycle in this parameter space is concerned, we can confine our interest to quadrupole fields of constant strength, say $|\mathbf{x}| = 1$. (This is justified

[1] Note that the basis of Γ matrices is not identical to the one used in Chapter 3, although there is a simple isomorphism.

[2] private communication Jan Budich

because the quadrupole energy is the only energy scale of the problem.) Note that the experimentally relevant scale of $|\mathbf{x}|$ defines the splitting between the two Kramers pairs and therefore the adiabatic operating frequencies of the proposed setup. Due to the mentioned SO(5) symmetry in the system [Avr+88], all possible quadrupole Hamiltonians $H(\mathcal{Q})$ are unitarily related by a Spin(5) representation of this SO(5) symmetry. The ten generators of this symmetry group of our family of Hamiltonians are given by [Avr+89]

$$\{V_i\}_i = \left\{ \frac{1}{2}\left[\Gamma_\alpha, \Gamma_\beta\right] = \Gamma_\alpha \Gamma_\beta \right\}_{\alpha<\beta}, \; i \in 0, \ldots, 9, \tag{8.8}$$

where $\alpha, \beta \in 0, \ldots, 4$. In order to show that indeed, the V_i generate a representation of SO(5), we spell out the orbit $t \mapsto H(t)$ of the group action in two different ways,

$$H(t) = \mathrm{e}^{t\frac{\hat{a}\mathbf{V}}{2}} x^\mu(0)\Gamma_\mu \mathrm{e}^{-t\frac{\hat{a}\mathbf{V}}{2}} = \left(\mathrm{e}^{t\hat{a}\mathbf{W}}\mathbf{x}(0)\right)^\mu \Gamma_\mu, \tag{8.9}$$

where $\hat{a}$ is a real ten-component unit vector specifying the direction of the SO(5) rotation. This shows the correspondence between the Spin(5) action on the space of Hamiltonians and the SO(5) action on the space $\mathbb{R}^5 \setminus \{0\}$ of parameter vectors $\mathbf{x} = x^\mu \mathbf{e}_\mu$. The generators of SO(5) are the real antisymmetric 5×5 matrices, and the vector $\mathbf{W}$ constitutes a basis of those. To prove (8.9) and explicitly calculate the basis $\mathbf{W}$ corresponding to $\mathbf{V}$, let us assume that $\hat{a}\mathbf{V} = \Gamma_\alpha\Gamma_\beta$ for some fixed $\alpha \neq \beta$. Then, $\hat{a}\mathbf{W} = \tilde{W}$ with a matrix $\tilde{W}$ having entries $\tilde{W}_{\alpha\beta} = -\tilde{W}_{\beta\alpha} = -1$ and zeros everywhere else, fulfils Eq. (8.9). To see this, we use the Baker-Hausdorff formula

$$e^{t\frac{\hat{a}\mathbf{V}}{2}} x^\mu \Gamma_\mu e^{-t\frac{\hat{a}\mathbf{V}}{2}} = \sum_{k=0}^{\infty} \frac{1}{k!}[x^\mu\Gamma_\mu, -\frac{t}{2}\Gamma_\alpha\Gamma_\beta]^{(k)}, \tag{8.10}$$

where the k-times commutator is defined like $[A,B]^{(k)} = [...[A,B],...,B]$. Note that the Spin(5) representation covers SO(5) twice. However, in (8.9), the Spin(5) group element $e^{t\frac{\hat{a}\mathbf{V}}{2}}$ appears twice, thus lifting the sign ambiguity. From the Clifford algebra properties, it follows that

$$\left(\tilde{W}\mathbf{x}\right)^\mu \Gamma_\mu = \frac{1}{2}[x^\mu\Gamma_\mu, -\Gamma_\alpha\Gamma_\beta] \tag{8.11}$$

and analogue for higher powers occurring in the exponential series, which shows the equality in (8.9). The $W_0, \ldots, W_3$ which correspond to the Spin(5) generators $V_0, \ldots, V_3$, that will be needed for explicit calculation of parameter loops, read

$$W_0 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \; W_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

$$W_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \; W_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{8.12}$$

We are interested in a cyclic time evolution $t \mapsto H(t)$ starting from $H(t=0) = \Gamma_0$ and given by a 2π SO(5) rotation on the space of quadrupole fields. As shown above, it is uniquely associated with a 2π Spin(5) rotation in Hilbert space, with the vector of generators given by $\mathbf{V}$. We call

$$P_0^{\pm} = \frac{1}{2}(1 \pm \Gamma_0) \tag{8.13}$$

the projector on the Kramers pair with eigenvalue $\pm|\mathbf{x}|$. In fact, due to our choice of the initial Hamiltonian, $P_0^{\pm}$ concurs with the projection on the HH/LH subspaces. Starting with a HH state $|\psi(0)\rangle$ satisfying $P_0^+|\psi(0)\rangle = |\psi(0)\rangle$, the adiabatic time evolution $U(t)$ can be conveniently expressed as [Sim83; Wil+84; Avr+89]

$$\begin{aligned} &U(t) = \lim_{n\to\infty} U_n(t) \text{ with} \\ &U_n(t) = P^+(t)P^+\left(\frac{(n-1)t}{n}\right)\cdots P^+\left(\frac{2t}{n}\right)P^+\left(\frac{t}{n}\right)P_0^+, \end{aligned} \tag{8.14}$$

where the time dependent projector on the Kramers pair with positive eigenvalue is given by $P^+(t) = \mathrm{e}^{t\frac{\hat{a}\mathbf{V}}{2}} P_0^+ \mathrm{e}^{-t\frac{\hat{a}\mathbf{V}}{2}}$. Along any such loop γ in parameter space the adiabatic evolution is readily computed analytically to yield [Avr+89]

$$U(t) = \mathrm{e}^{t\frac{\hat{a}\mathbf{V}}{2}} \mathrm{e}^{-tP_0^+\frac{\hat{a}\mathbf{V}}{2}P_0^+}. \tag{8.15}$$

We can check this relation by showing that it fulfils the desired boundary condition $P^+(t)U(t) = U(t)P_0^+$ and the correct differential equation $\frac{d}{dt}U(t)P_0^+ = -\mathcal{A}UP_0^+$, where $\mathcal{A}$ is the generator of adiabatic evolution as defined in (8.2). Explicitly, we find that

$$\frac{d}{dt}U(t)P_0^+ = e^{t\frac{\hat{a}\mathbf{V}}{2}}\left(1-P_0^+\right)\frac{\hat{a}\mathbf{V}}{2}e^{-tP_0^+\frac{\hat{a}\mathbf{V}}{2}P_0^+}P_0^+ = \left(1-P^+(t)\right)\frac{\hat{a}\mathbf{V}}{2}U(t)P_0^+ = -\mathcal{A}U(t)P_0^+. \tag{8.16}$$

The first factor in (8.15) gives $\mathrm{e}^{2\pi\frac{\hat{a}\mathbf{V}}{2}} = -1$ once the loop is completed. The second factor at $t = 2\pi$ defines an SU(2) transformation on the HH subspace which is the desired holonomy U_γ (see Eq. (8.3)) up to a sign. Note that the holonomy associated with a loop γ is a purely geometrical object. It does not depend on parameterization, i.e. on the time-dependent velocity with which the electric field is ramped, as long as the adiabatic approximation is justified.
We now explicitly construct the direction $\hat{a}$ needed to obtain any holonomy as parametrized in Eq. (8.1). The angle and axis of the rotation can be tuned using the relations

$$P_0^+\Gamma_0\Gamma_\mu P_0^+ = 0,\ \mu \neq 0, \tag{8.17}$$

as well as

$$\begin{aligned} &P_0^+\Gamma_4\Gamma_1P_0^+ = i\sigma_x,\ P_0^+\Gamma_1\Gamma_3P_0^+ = i\sigma_y, \\ &P_0^+\Gamma_1\Gamma_2P_0^+ = i\sigma_z, \end{aligned} \tag{8.18}$$

where σ_i are the Pauli matrices on the HH subspace with eigenvalue $+|\mathbf{x}|$. To see this, let us restrict ourselves to the four generators $\Gamma_1\Gamma_\mu,\ \mu \neq 1$ and label them $V_0 = \Gamma_0\Gamma_1$, $V_1 =$

$\Gamma_4\Gamma_1$, $V_2 = \Gamma_1\Gamma_3$, $V_3 = \Gamma_1\Gamma_2$. With this restriction $\hat{a}$ only has the nonvanishing components a_0, a_1, a_2, a_3 satisfying $\sum_{i=0}^{3} a_i^2 = 1$. Using Eqs. (8.15-8.18) we get by comparison to Eq. (8.1)

$$\varphi = 2\pi \left(1 - \sqrt{\sum_{i\neq 0} a_i^2}\right) = 2\pi \left(1 - \sqrt{1 - a_0^2}\right) \in [0, 2\pi]\,,$$
$$\hat{n} = \frac{(a_1, a_2, a_3)}{|(a_1, a_2, a_3)|}. \tag{8.19}$$

Next, we translate the loop associated with the direction $\hat{a}$ into a time dependent quadrupole field, using the relation (8.9) and SO(5) generators $W_0, ..., W_3$ of (8.12). Now, we can define the time dependent quadrupole field associated with the loop in direction $\hat{a}$:

$$\mathcal{Q}(t) = x^\mu(t) Q_\mu = \left(\mathrm{e}^{t\hat{a}\mathbf{W}}\mathbf{x}(0)\right)^\mu Q_\mu, \quad t \in [0, 2\pi] \tag{8.20}$$

which needs to be experimentally applied to perform the desired single qubit operation.

Let us give a concrete example. If we were to rotate the HH spin from pointing in z-direction to the x-direction this would correspond to the operation $\mathcal{U}(-\hat{e}_y, \frac{\pi}{2}) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ which is associated with the quadrupole field

$$Q(t) = \left(\mathrm{e}^{t\left(\frac{\sqrt{7}}{4}W_0 - \frac{3}{4}W_2\right)}\mathbf{e_0}\right)^\mu Q_\mu, \quad t \in [0, 2\pi]\,, \tag{8.21}$$

i.e. $\hat{a} = (a_0, a_1, a_2, a_3) = (\frac{\sqrt{7}}{4}, 0, -\frac{3}{4}, 0)$ and $\mathbf{x}(t=0) = \mathbf{e_0} = (1, 0, 0, 0, 0)$ in the language of our general analysis. Indeed plugging this choice of $\hat{a}$ into Eq. (8.19) yields $\hat{n} = -\hat{e}_y$, $\varphi = \frac{\pi}{2}$. A stroboscopic illustration of a possible electrostatic gating scheme realizing this time-dependent quadrupole field is shown in Fig. 8.2. For this particular example, we only need 10 of the 18 gates illustrated in Fig. 8.1. To perform an arbitrary SU(2) transformation 14 of these 18 gates are needed. We could drop, for instance, the four gates that are coloured in red and green in Fig. 8.1 and still be able to perform any desired single qubit rotation on the HH subspace.

8.1 Estimation of experimental parameters in GaAs quantum dots

Up to now, the energy scale $\Delta E = |\mathcal{Q}| = |\mathbf{x}|$ (see Eq. (8.5)) has been treated as a free parameter. To show that this scale is amendable to state of the art experiments on GaAs quantum dots, we give a numerical estimate for ΔE. To do so, we calculate the HH-LH splitting ΔE associated with an electrostatic potential $e\Phi_4(\mathbf{r}) = \lambda \mathbf{r}^T \mathcal{Q} \mathbf{r}$ with quadrupole symmetry on the basis of a Luttinger four-band model for the valence bands of a GaAs/AlGaAs quantum well [And+87; Chu91]. Here, $\mathbf{r}$ denotes the real space position vector and the QDs are modelled by a parabolic lateral confinement potential defining the dots on a typical length scale of 50nm. The strength of the potential is determined by the constant λ. For a realistic quadrupole potential $e\Phi_4 \sim$ 50 meV at a distance $r \sim 50$ nm away from the center of the dot, we obtain a splitting of $\Delta E = 0.57$ meV, which corresponds to a temperature of 6.6 K and an adiabatic frequency of $\omega = 0.87$ THz respectively. Therefore, it is easily possible to stay below this frequency such that the adiabatic evolution is justified and at the same time complete the loop much faster

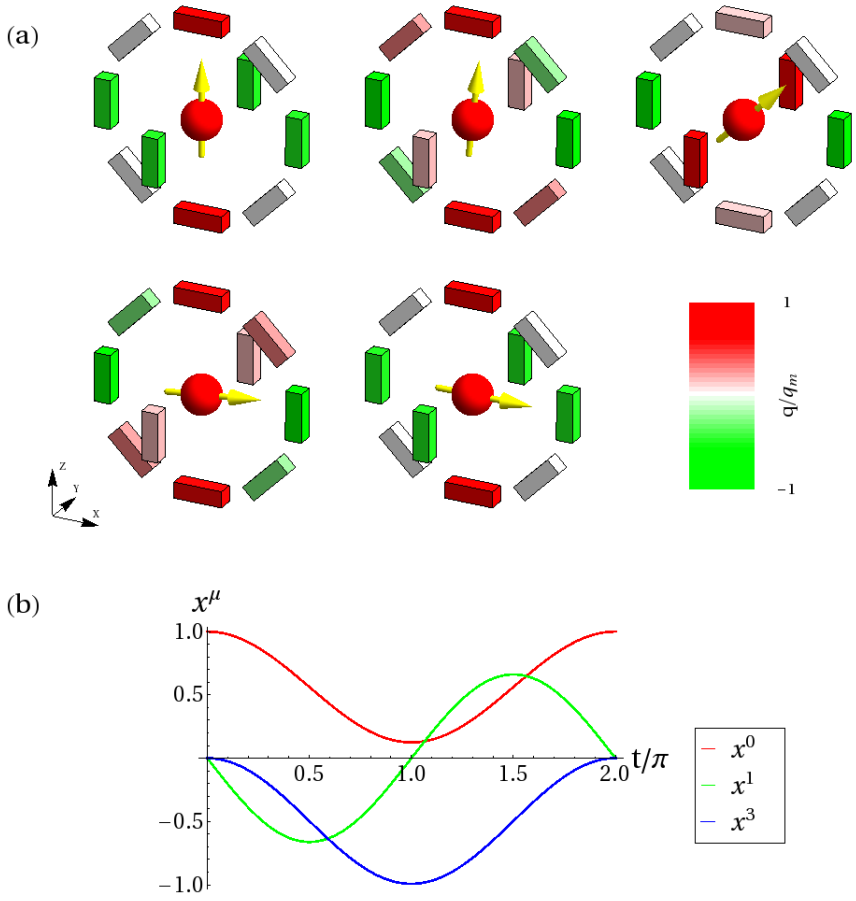


Figure 8.2: (a) A 10-gate setup realizing the operation $\mathcal{U}(-\hat{e}_y, \frac{\pi}{2})$ on the HH spin (yellow arrow). The colors of the schematic gates visualize their time-dependent charge during the loop operation, at times from left to right and top to bottom, $t = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$, and 2π. All charges are normalized to the charge q_m of the topmost gate at $t = 0$. (b) Time-dependence of non-zero components of $\mathbf{x}$ during the operation $\mathcal{U}(-\hat{e}_y, \frac{\pi}{2})$. Taken from [Bud+12a]. © (2012) by the American Physical Society

than typical dephasing times in HH spin qubits. (T_2 of the order of μs has been measured in Ref. [deGre+11].)

In real experiments, there will not only be the (wanted) HH-LH splitting ΔE induced by the quadrupole field but also an (unwanted) HH-LH splitting ΔE_0 induced by confinement. For our purposes, the former should be much larger than the latter. We estimate in Section 8.1.1 that often times it is the other way round, i.e. ΔE_0 is much larger than ΔE which is a true problem for our proposal. However, by applying mechanical strain, the splitting of the individual quadruplet subbands of the quantum dot can be engineered significantly [And+87; Chu95]. For the parameters used in our model, the confinement induced splitting ΔE_0 can then be realistically tuned below our estimated value of $\Delta E = 0.57$ meV. Hence, strain engineering of the QD is needed to guarantee a reliable performance of our setup. Additionally, we note that our proposal is robust against unwanted residual dipole fields, deviations from a quadrupole potential with only $l = 2$ contributions, and deviations from a quadratic confinement potential. The influence of these perturbations on ΔE are carefully analysed in Section 8.1.2 and shown to be harmless.

8.1.1 Quadrupole induced HH/LH splitting in strained GaAs quantum dots

In this section, we give a quantitative estimate of the heavy hole (HH)/light hole (LH) splitting induced by an electric quadrupole field on strained GaAs quantum dots. We model a quantum dot using the effective 2D Hamiltonian of a [001] quantum well (QW) [And+87] and add some parabolic confinement $\Phi_1(x, y)$. This reduces the symmetry to D_{2d} and therefore, even without a quadrupole potential, we expect a HH/LH splitting ΔE_0. We account for this by extending the Hamiltonian $H(\mathcal{Q}) = \mathbf{J}^T \mathcal{Q} \mathbf{J}$ to

$$H' = H(\mathcal{Q}) + \frac{\Delta E_0}{2} \tau_z \tag{8.22}$$

with $\tau_z = \mathrm{diag}(1, -1, -1, 1)$, and the Hamiltonian is written in the basis

$$\left\{ \left| \frac{3}{2}, \frac{3}{2} \right\rangle, \left| \frac{3}{2}, \frac{1}{2} \right\rangle, \left| \frac{3}{2}, -\frac{1}{2} \right\rangle, \left| \frac{3}{2}, -\frac{3}{2} \right\rangle \right\}. \tag{8.23}$$

Without loss of generality, we use a quadrupole potential $e\Phi_4 = \frac{1}{R^2} \mathbf{r}^T \mathcal{Q} \mathbf{r}$ associated with the quadrupole tensor of four Coulomb charges $\pm q$ at equal radius R in the (x, y) plane (corresponding to $\lambda = \frac{1}{R^2}$),

$$\mathcal{Q} = \frac{1}{4\pi\epsilon} \frac{6eq}{R} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{8.24}$$

Whereas the spectrum of H is $E = \pm|\mathbf{x}|$, where $\mathbf{x}$ is a 5-component vector defined by the expansion $\mathcal{Q} = x^\mu Q_\mu$, the spectrum of H' simplifies for our choice of the quadrupole potential to

$$E = \pm \frac{1}{2} \sqrt{\Delta E_0^2 + 4|\mathbf{x}|^2}. \tag{8.25}$$

We will use this relation to fit $|\mathbf{x}|$ as a function of the strength of the quadrupole potential. Note that here, $\mathcal{Q} \propto Q_1$, and we could also model the zero-field splitting ΔE_0 by the quadrupole Hamiltonian $\frac{\Delta E_0}{2} \mathbf{J}^T Q_0 \mathbf{J}$. This shows that one could try to use a voltage offset applied to the

quadrupole gates, in order to get rid of the undesired splitting ΔE_0. However, it turns out that the required voltage offset would be so large that it would destroy the in-plane confinement of the QD. Therefore, we choose a different approach below, by applying mechanical strain to tune ΔE_0 to zero.

To obtain an effective Hamiltonian for the QW, we first solve the envelope function $\mathbf{f}(z)$ where z is the direction of growth. In general, the 4-component envelope function $\mathbf{f}(z)$ depends on $\mathbf{k}_\parallel = (k_x, k_y)$. We simplify the problem by performing a $\mathbf{k}\cdot\mathbf{p}$ calculation with expansion of $\mathbf{k}_\parallel$ around the Γ point. The Luttinger Hamiltonian $H_L(\mathbf{k}_\parallel = 0)$ is diagonal and for the ith component f_i of $\mathbf{f}$ we find

$$\left(k_z \frac{1}{2m_i(z)} k_z + V(z)\right) f_i(z) = E_i f_i(z). \tag{8.26}$$

Here, $m_i(z)$ is the material dependent bulk effective mass, which is $m_{B,i}$ for the barrier and $m_{W,i}$ for the material of the well and band dependent (index i). Furthermore, $V(z) = V_B$ in the barrier and zero otherwise. We use the symmetric ansatz

$$f_i(z) = \begin{cases} A_i e^{\xi_i (z+W/2)} & z < -W/2, \\ B_i \cos(k_i z) & -W/2 \le z \le W/2, \\ A_i e^{-\xi_i (z-W/2)} & z > W/2, \end{cases} \tag{8.27}$$

where $W = 60\text{nm}$ is the QW width, $k_i = \sqrt{2 m_{W,i} E_i}$ and $\xi_i = \sqrt{2 m_{B,i}(V_B - E_i)}$. Continuity of $f_i(z)$ and $f_i'(z)/m_i(z)$ give the secular equation[3]

$$\sqrt{\frac{1}{\tilde{k_i}^2} - 1} = \left(\frac{m_{B,i}}{m_{W,i}}\right)^{1/2} \tan\left(\tilde{k_i} W \sqrt{\frac{m_{W,i} V_B}{2}}\right) \tag{8.28}$$

with $\tilde{k_i}\sqrt{2 m_{W,i} V_B} = k_i$.

The Luttinger Hamiltonian for Γ_8 bands including corrections due to strain reads

$$H_L = -\begin{pmatrix} P+Q & -S & R & 0 \\ -S^\dagger & P-Q & 0 & R \\ R^\dagger & 0 & P-Q & S \\ 0 & R^\dagger & S^\dagger & P+Q \end{pmatrix} \tag{8.29}$$

written in the basis (8.23). The strain tensor ϵ_{ij} gives the displacement of an atom at unit vector $\hat{i}$ along unit vector $\hat{j}$. We consider only uniaxial strain with $\epsilon_{xx} = \epsilon_{yy} \neq \epsilon_{zz}$ and $\epsilon_{xy} = \epsilon_{xz} = \epsilon_{yz} = 0$. Then, only P and Q include corrections due to strain:

$$P = t_0 \gamma_1 (k_x^2 + k_y^2) + t_0 k_z \gamma_1 k_z + P_\epsilon, \tag{8.30}$$

$$Q = t_0 \gamma_2 (k_x^2 + k_y^2) - 2 t_0 k_z \gamma_2 k_z + Q_\epsilon, \tag{8.31}$$

$$R = t_0 \sqrt{3}(-\gamma_2 (k_x^2 - k_y^2) + 2 i \gamma_3 k_x k_y), \tag{8.32}$$

$$S = t_0 \sqrt{3}(k_x - i k_y)\{\gamma_3, k_z\} \tag{8.33}$$

[3] Unfortunately, in our paper [Bud+12a] there is a mistake at this point, which we have noticed only after publication. There, we incorrectly wrote $\left(\frac{m_{B,i}}{m_{W,i}}\right)^{3/2}$ instead of $\left(\frac{m_{B,i}}{m_{W,i}}\right)^{1/2}$ in the secular equation. However, redoing the calculations shows that both the quadrupole induced HH/LH splitting $|\mathbf{x}|$ and the required applied strain ζ change by less than 1%, so all the conclusions of the paper remain valid. The smallness of the change is because the effective mass factor only affects the matching condition of the wave function at the barrier/well interface, and thereby mainly changes the wave function in the region where it decays exponentially. The figures shown here are those of our publication.

with $t_0 = \frac{1}{2m_0}$ and

$$P_\epsilon = -a_v(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}), \tag{8.34}$$

$$Q_\epsilon = -\frac{b}{2}(\epsilon_{xx} + \epsilon_{yy} - 2\epsilon_{zz}). \tag{8.35}$$

The GaAs/AlAs lattice constants are almost the same (5.65 Å vs. 5.66 Å). This is desirable because one needs rather wide quantum wells and intends to avoid uncontrolled relaxation of strain. Here, we assume additional strain due to external pressure τ_{zz} which can be expressed in terms of the stiffness tensor C relating strain and stress. The condition of no transversal stress $\tau_{xx} = \tau_{yy} = 0$ gives

$$\epsilon_{xx} = \epsilon_{yy} = \frac{-C_{12}}{C_{11}^2 + C_{11}C_{12} - 2C_{12}^2}\tau_{zz} \tag{8.36}$$

$$\epsilon_{zz} = \frac{C_{11} + C_{12}}{C_{11}^2 + C_{11}C_{12} - 2C_{12}^2}\tau_{zz}, \tag{8.37}$$

where $C_{11} = 11.88\ 10^5$ bar and $C_{12} = 5.38\ 10^5$ bar [Chu95]. We take the same values for barrier and QW for the deformation potentials, $a_v = 1.16$ eV and $b = -1.7$ eV. [Chu95] The parameter $\zeta := Q_\epsilon$ will be used as strain control. A pressure of 1kbar gives $\zeta = 2.61$ meV. P_ϵ is an unimportant overall energy shift.

The effective QW Hamiltonian is obtained by integration over envelope functions $f_\alpha(z)$ of the lowest LH and HH QW subbands,

$$H_{\alpha\beta}^{QW} = \int \mathrm{d}z\, f_\alpha^\dagger(z) H_L f_\beta(z). \tag{8.38}$$

Contributions of higher subbands give quantitative, but not qualitative changes of our estimates, since strain gives a diagonal correction to H^{QW} and can be used to tune ΔE_0. Together with the in-plane potentials Φ_i, H^{QW} gives our QD model which is numerically diagonalized.

For a quantitative estimate of $|\mathbf{x}|$, we use the same parameters as Ref. [And+87]: For GaAs, $\gamma_1 = 6.85$, $\gamma_2 = 2.1$, $\gamma_3 = 2.9$. For AlAs, $\gamma_1 = 3.45$, $\gamma_2 = 0.68$, $\gamma_3 = 1.29$. The barrier material is $Al_{1-\nu}Ga_\nu As$ with $\nu = 0.21$ and the Luttinger parameters are obtained by linear interpolation. The bulk gap difference is $\Delta E_g = (1.04\nu + 0.47\nu^2)$ eV $= 0.239$ eV. We assume that the valence band shift from well to barrier is $-0.4\ \Delta E_g$.

For the in-plane confinement, we use $e\Phi_1 = -0.15$ eV$(r/R_{\max})^2$ where $r^2 = x^2 + y^2$. $e\Phi_1(R_{\max})$ should not exceed ΔE_g. Φ_1 is discretized on a lattice corresponding to L=100 nm side length, so $R_{\max} = 50$ nm. By choosing $R_{\max}$ and W comparable, we intend to have about the same level spacing due to in-plane and QW confinement. Then, the confinement comes closer to the ideal, fully rotationally symmetric confinement.

With this geometry, a value of $e\Phi_i(R_{\max}) = -1$ eV gives a field strength of 40 meV/nm at $R_{\max}$. Fig. 8.3 shows the zero-field splitting ΔE_0 as function of strain, demonstrating that the confinement induced splitting can be tuned down to zero by means of uniaxial strain. Fig. 8.4 shows fits to the dispersion (8.25) in order to obtain the quadrupole induced splitting $2|\mathbf{x}|$. A realistic quadrupole with a maximum potential $e\Phi_4|_{r=50\text{ nm}}$ of 50 meV gives a quadrupole induced splitting of $2|\mathbf{x}| \approx 0.57$ meV.

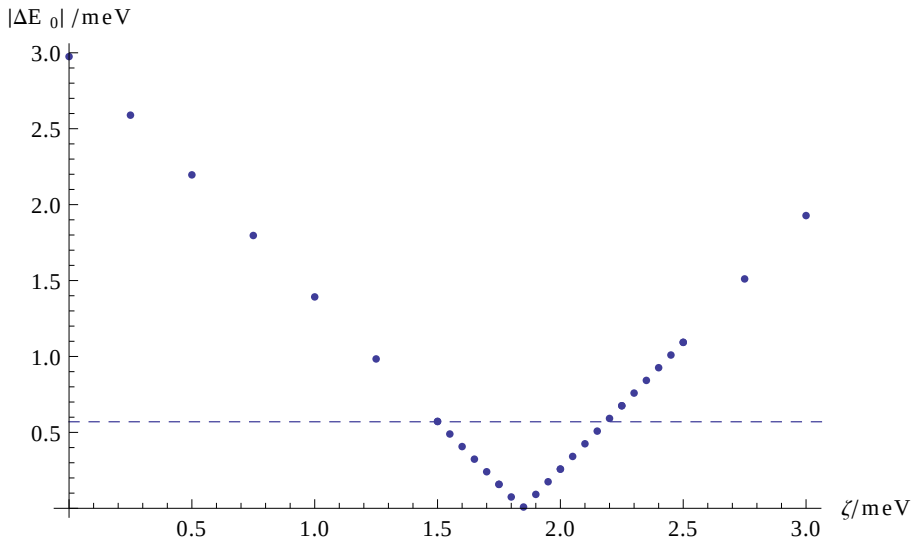


Figure 8.3: HH/LH splitting ΔE_0 (in the absence of a quadrupole field) as a function of the strain-induced subband shift ζ for a QW thickness $W = 60$ nm. Evidently, the (unwanted) HH/LH splitting ΔE_0 can be tuned down to zero by a uniform strain in z direction. The dashed line marks the value of the typical (wanted) HH/LH splitting $\Delta E = 0.57$ meV due to a quadrupole field as discussed in the text. Taken from [Bud+12a]. © (2012) by the American Physical Society

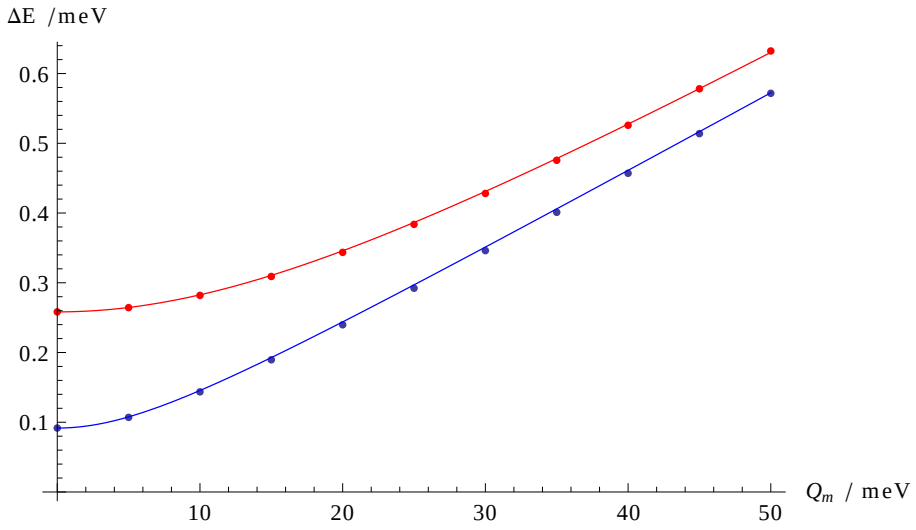


Figure 8.4: HH/LH splitting ΔE as function of the quadrupole potential $Q_m = \max(e\Phi_4)$ at $r = 50$ nm. The dots are numerical results and the continuous lines fits to the expected dispersion (8.25) with the quadrupole parameter $|\mathbf{x}| = 0.00575\ Q_m$ (red) and $|\mathbf{x}| = 0.00565\ Q_m$ (blue). The full red line corresponds to a strain energy $\zeta = 2$ meV and the full blue line to $\zeta = 1.9$ meV. Taken from [Bud+12a]. © (2012) by the American Physical Society

8.1.2 Stability of the quantum dot setup against perturbating potentials

The aim of this section is to analyse the stability of the effective quadrupole Hamiltonian H' against deviations from a perfect quadrupole potential with $l = 2$. These deviations include external dipole fields and deviations from the quadratic confinement and will be described as $V(r, \phi)$ in the following. The stability of H' implies the stability of the quadrupole Hamiltonian $H(\mathcal{Q})$ since a change in the unwanted ΔE_0 can be suppressed by adjusting the strain.

We consider the axial multipole expansion of the in-plane potential $V(r, \phi)$ given by a distribution of Coulomb charges $\rho(R, \phi')$. The QD extension is small against the distance to the gates, i.e. $r \ll R$. We expand in the Legendre Polynomials P_l,

$$V(r,\phi) = \frac{e}{4\pi\epsilon}\sum_{l=0}^{\infty} r^l \int_0^{2\pi} d\phi' P_l(\cos(\phi-\phi')) \int_0^{\infty} dR \frac{1}{R^l}\rho(R,\phi'). \tag{8.39}$$

We continue by expanding the P_l as

$$r^l P_l(\cos(\phi-\phi')) = r^l \sum_{j=l,l-2,..} \alpha_{l,j}\cos(j(\phi-\phi')). \tag{8.40}$$

For the quadrupole symmetry $V(r, \phi + \frac{\pi}{2}) = -V(r, \phi)$ and upon inserting (8.40) into (8.39), the nonzero coefficients $\alpha_{l,j}$ have $j = 2, 6, 10, \ldots$ and $j \leq l$. Similarly, for the dipole symmetry $V(r, \phi + \pi) = -V(r, \phi)$, the non-zero coefficients $\alpha_{l,j}$ fulfil $j = 1, 3, 5, \ldots$ and $j \leq l$. Table 8.1 shows how some characteristic terms in the expansion (8.39) enter our model.

$l = 0$		Overall shift in energy that does not change ΔE.
$l = 1$	$r\cos\phi$	Shift of the center of the bound state assuming that quadrupole and confining potentials ($\Phi_1 + \Phi_4$) are quadratic in r. ΔE unchanged.
$l = 2$	r^2, $r^2\cos 2\phi$	Included in the model as $\Phi_1 + \Phi_4$.
$l = 3$	$r^3 P_3 = r^3(\frac{3}{8}\cos\phi + \frac{5}{8}\cos 3\phi)$	Lowest order that appears in dipole expansion and can induce quadratic Stark effect.
$l = 4$	$r^4\cos 4\phi$	Deviation from quadrupole symmetry by four equally charged gates.
	$r^4\cos 2\phi$	Allowed by quadrupole symmetry leading to the same effective Hamiltonian $H(\mathcal{Q})$ with $J = \frac{3}{2}$ but with the induced value ΔE only a few percent in comparison with $l = 2$ term. Does not influence holonomy operations.
	r^4	Correction to the confinement potential, which removes stability against the $l = 1$ perturbation.
$l = 6$	$r^6\cos 6\phi$	Lowest order perturbation that appears in quadrupole expansion.

Table 8.1: Characterization of terms in the axial multipole expansion (8.40) by their ability to perturb the holonomy.

Let us now summarize the results included in Table 8.1. The $l = 0$ term induces an uninteresting energy shift. The $l = 1$ term could give rise to a linear or quadratic Stark effect. However,

in very good approximation, we may assume that GaAs and AlAs have inversion symmetry and can be described by a Luttinger Hamiltonian. Since the Luttinger Hamiltonian H_L is even under inversion, the lowest bound states have even parity. This excludes the linear Stark effect by symmetry. Further, as long as we model both the confinement and the quadrupole potential as quadratic in r, a linear potential will simply shift the center of the wave function. Thus, the quadratic Stark effect also cannot change ΔE.

For a numerical estimate of higher-l terms, we model the gates by four Coulomb charges at $r = 50$ nm. We find that the $l = 3$ and $l = 4$ terms barely change ΔE even if the corresponding charge imbalance at the gates is highly overestimated as compared to a realistic experimental setup, meaning we have chosen them of the order of the quadrupole charges itself. If quadrupole symmetry of the potential holds, the lowest perturbation term is $l = 6$. This term will change depending on the shape of the gates, but, since it contains a small parameter (r^6/R^6), it is negligible.

Finally, we note that the system is no longer robust against the quadratic Stark effect if the confinement potential behaves other than r^2. We analyze this case in Fig. 8.5, by changing the in-plane confinement to $e\Phi_1' = -0.15\ \mathrm{eV}\frac{r^2}{R_{\max}^2}(1 + \frac{1}{3}\frac{r^2}{R_{\max}^2})$. A residual constant dipole field is modelled by an additional potential $e\Phi_2 = -0.025\ \mathrm{eV}\frac{r}{R_{\max}}\cos(\phi - \frac{\pi}{3})$ so that it is not aligned with the other potentials, and corresponds to a dipole charging being $\frac{1}{3}$ of the quadrupole charging, assuming the latter to yield a maximum potential $e\Phi_4|_{r=50\mathrm{nm}}$ of 50 meV. This certainly overestimates the error expected in an experiment. Nevertheless, as can be seen in Fig. 8.5, ΔE is barely affected by this perturbation. In Fig. 8.5, the dipole potential is kept constant for the red dots, while the quadrupole potential scales with Q_m. The blue dots show ΔE without the dipole potential, for comparison. Although not apparent from the figure, we have checked that the difference given by red vs. blue dots scales with the square of the dipole potential.

Summarizing, we find that the effective Hamiltonian $H(\mathcal{Q})$ remains valid in good approximation. In all cases, the quadrupole splitting dominates the other (disturbing) contributions for realistic parameters.

8.2 Summary and outlook

In summary, we have demonstrated that an electric quadrupole field can be used to fully control a HH qubit without breaking TRS. The adiabatic time scale of our proposal is determined by the field induced splitting ΔE between the two Kramers pairs, which we have estimated for GaAs QDs to be on the order of 0.57 meV. The maximum operating frequency of the device should be significantly below this energy scale to justify the adiabatic approximation which is understood throughout our analysis. Confinement induced splitting between the two Kramers pairs in the $J = \frac{3}{2}$ quadruplet of levels at the relevant energy in the HH QD impinges on the efficiency of the geometric control over the qubit. The scale of this splitting for a given quadruplet can be tuned/reduced by applying strain. We note that exact control over the qubit is still possible as long as the quadrupole energy gap is larger than the confinement induced splitting. Our proposal is not limited to HH quantum dots in GaAs quantum wells, but can in principle also be employed to process trapped spin-$\frac{3}{2}$ ions or HH-like valence band impurities by means of a quadrupole field. The presence of TRS in combination with suppressed hyperfine coupling in HH systems renders our proposal less prone to decoherence than non-adiabatic processing schemes relying on the presence of a Zeeman splitting due to an external magnetic field. Two-qubit gates can be performed by virtue of electrostatic gates as proposed in Ref. [Los+98]. All-electric spin pumping

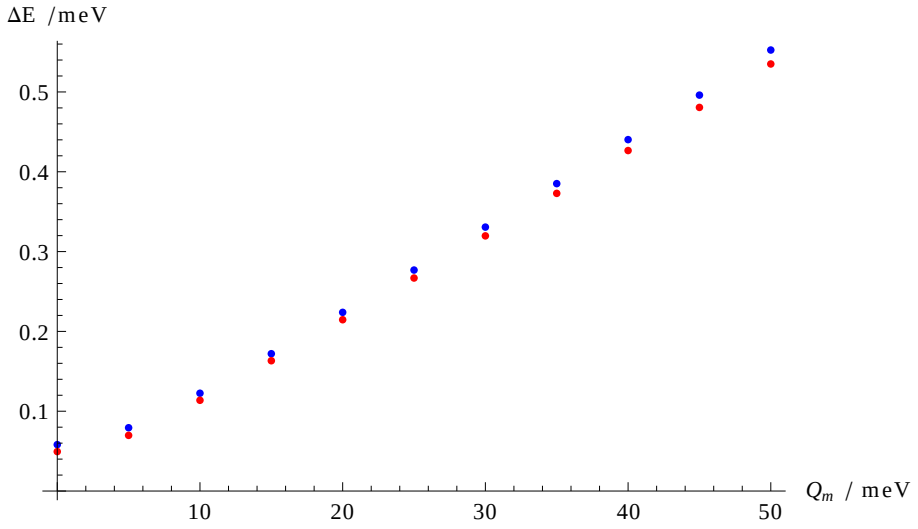


Figure 8.5: Including a r^4 correction to the confinement (Φ_1' in the text) allows for the quadratic Stark effect by a homogeneous electric field. The plot shows the HH/LH splitting ΔE as function of the quadrupole potential $Q_m = \max(\mathrm{e}\Phi_4)$ at $r = 50$ nm with $\zeta = 1.9$ meV and $W = 60$ nm. Blue dots are without the dipole potential Φ_2 while red dots include the Φ_2, which corresponds to a charging ratio of 1/3 of a dipole vs. quadruple configuration at $Q_m = 50$ meV. This ratio certainly overestimates the error that we expect in the experimental situation. Taken from [Bud+12a]. © (2012) by the American Physical Society

and spin filtering techniques respectively [Bro+10b] can be used to perform initialization and readout tasks on the quantum dots. Hence, our proposal in principle allows for TRS preserving universal quantum computing.

Chapter 9

Spin-dependent thermoelectric transport in HgTe/CdTe quantum wells

In this chapter, we analyse thermally induced spin and charge transport in HgTe/CdTe quantum wells on the basis of the numerical non-equilibrium Green's function technique in the linear response regime. In the topologically non-trivial regime, we find a clear signature of the gap of the edge states due to their finite overlap from opposite sample boundaries – both in the charge Seebeck and spin Nernst signal. We are able to fully understand the physical origin of the thermoelectric transport signatures of edge and bulk states based on simple analytical models. We find that the spin Nernst signal is related to the spin Hall conductance by a Mott-like relation which is exact to all orders in the temperature difference between the warm and the cold reservoir. The theoretical foundations needed for this analysis have been discussed in Chapter 5 on electronic transport and Chapters 6 and 7 on the two-dimensional topological insulator system of HgTe/CdTe quantum wells. This chapter follows closely our publication [Rot+12].

9.1 Introduction and overview

Thermoelectric transport coefficients define the efficiency of a system to generate an electrical power from a temperature gradient [Mac62]. The most established thermoelectric phenomenon is the Seebeck effect [See26], in which a current (closed boundary conditions), or a bias (open boundary conditions) is induced from a temperature difference held between two reservoirs of a junction. The transverse Seebeck coefficient, or Nernst coefficient, refers to the alternative situation where the thermally induced current (bias) flows in the direction transverse to both the temperature gradient and the applied magnetic field [Ner87].

Thermoelectric effects have major consequences in terms of technological impact and scientific understanding. On the one hand, these effects offer interesting applications based on heat-voltage conversion: thermometry, refrigeration, power generation [Gia+06; Bel08]. On the other hand, thermoelectric coefficients combine information from energy and charge flows at quasi-equilibrium. Furthermore, they are more sensitive to the details of the density of states than electrical conductance [Abr88; Zim60; Bee+92]. Both aspects make them a powerful tool to probe the system dynamics [Seg05].

During the last two decades, there have been considerable technological advances in low-temperature nanoscale physics. This allows precise measurements of thermoelectric transport signals, obtained in various systems like bismuth [Beh+07], superconductors [Bel+04; Cha+10],

carbon-based structures [Bal11], or molecular junctions [Dub+11].

The recent alliance of spintronics and thermoelectric transport brings up a spin analogue of Seebeck and Nernst effects (see Ref. [Bau11] for a short review). Especially in systems with strong spin-orbit interactions, a temperature gradient can generate a transverse spin current (or bias) even in the absence of an applied magnetic field. This can lead to the anomalous Nernst effect (in the case of ferromagnetic systems)[Miy+07; Sla+11; Han+08a] or the spin Nernst effect (in the time-reversal symmetric situation)[Chu+10; Liu+10c; Dyr+12].

Systems with strong spin-orbit interactions have been extensively studied in condensed matter physics especially since the prediction of the spin Hall effect [Dya+71; Hir99; Mur+03; Sin+04] which allows for an all-electrical manipulation of spin. The spin Hall effect generates a transverse spin accumulation as a response to a longitudinal applied electric field. Spin-orbit interactions have several origins which distinguish the different types of phenomena, for instance, an *extrinsic* spin Hall effect can emerge from the spin-orbit dependent scattering on impurities or defects [Dya+71; Hir99; Han+06; Han+09a; Kat+04; Gar+10]. On the other hand, bulk or structure inversion asymmetries [Byc+84; Dre55; Win05] give rise to an *intrinsic* spin Hall effect [Mur+03; Sin+04; Brü+10], which may be described in terms of an anomalous velocity or a spin-dependent classical force [Ber84; Xia+10; Sun+99], compare also Chapter 4.2.

Recent experiments have demonstrated the existence of an intrinsic spin Hall effect in HgTe/CdTe quantum wells (QWs) [Brü+12] by the use of the quantum spin Hall effect as the detector. This novel electronic phase is characterized by an insulating bulk and protected metallic edge states. The emergence of the quantum spin Hall effect is due to strong spin-orbit coupling and other relativistic corrections, which reverse bands of opposite parities. The electrons obey a massive Dirac equation and the sign of the mass term enables us to distinguish the topological phases. The edge states consist of Kramers pairs moving in opposite direction at each boundary [Kan+05a; Fu+07a; Ber+06a] and time-reversal symmetry protects them from non-magnetic and elastic backscattering [Wu+06]. Thereby, these edge channels carry "dissipation-less" spin currents whose existence in HgTe QWs has been confirmed experimentally by measurements in multi-terminal devices [Kön+08].

A major goal in research on thermoelectrics is increasing the efficiency of power conversion, for harvesting electrical power from heat. Therefore, much attention is often given to the figure of merit $ZT = \sigma S^2 T/\kappa$. ZT is a measure of the electrical power that can be harvested, compared to the used heat power. Here σ, S, κ and T are electrical conductivity, the thermopower (Seebeck coefficient), total heat conductivity including electron and phonon contributions, and the temperature, respectively. ZT can be calculated from the Onsager reciprocal relations, which relate electrical and heat currents to voltage and temperature gradients. In order to maximize ZT, one must minimize thermal conductivity - in particular, the phonon contribution increases thermal conductivity but not electrical conductivity or thermopower, and is detrimental. At the same time, one should maximize thermopower S and electrical conductivity - which is a difficult task since electrical and thermal conductivity are related, e.g. in metals by the Wiedemann-Franz law $\frac{\kappa}{\sigma} = \frac{\pi^2 k_B^2}{3e^2} T$. High thermopower requires steep dependence of the electronic density of states on the energy.[1] While this is difficult to achieve in metals, semiconductors are good candidates since engineering (by doping or gating) allows for choosing the Fermi energy just above the bottom of the conduction band. Good candidates for thermoelectric materials that fulfil these

[1] Phrased differently, a thermoelectric signal is related to the particle-hole asymmetry of the energy dispersion, with respect to the chemical potential. For this reason, superconductors show vanishing thermoelectric signals and are useful reference materials in thermocouples with other materials of interest.

combined requirements are semiconductors consisting of heavy atoms - typically they have a narrow (or even inverted) gap, which helps increasing electrical conductivity. Interestingly, this is just the class of materials that also turn out to be (3D) topological insulators, e.g. Bi_2Te_3 [Fu+07b; Zha+09]. Therefore, topological insulators have also attracted the interest of the scientific community working on thermoelectricity. Several proposals have been conceived how to modify the bulk of a topological insulator in a way to decrease phonon contributions, and at the same time benefit from the high conductivity of surface or edge states. The latter could be line dislocations [Tre+10], surfaces states of a 3D TI [Tre+11], or edge states of a narrow ribbon of 2D TI [Tak+10]. The aim is the enhancement of the contribution of edge states to the thermoelectric transport compared to the bulk modes. However, the subject of this thesis is not thermoelectric power conversion but spintronics, and we will not further discuss efficiency of power conversion and thermal conductivity. In this chapter we are rather interested in thermoelectricity as a tool acting on the spin of a 2D topological insulator.

In this work, we investigate the spin-dependent thermoelectric transport in topological insulators based on HgTe/CdTe QWs, in the absence of magnetic fields. The behaviour of the Seebeck coefficient and the spin Nernst signal is analysed in a four-terminal cross-bar setup, as shown in Fig. 9.1. A thermal gradient between lateral leads induces a longitudinal electrical bias and a transverse spin current. Each of them can be used as a probe of the topological regime as well as finite size effects of the quantum spin Hall insulator. We show that the oscillatory character of the Seebeck and spin Nernst coefficients in the bulk gap highlights the presence of the mini-gap – due to the finite overlap of the edge states from opposite sample boundaries. Furthermore, we describe a qualitative relation between the type of particles in a given band and the magnitude of the spin Nernst signal. This allows us to provide a natural explanation of the observed phenomena based on anomalous velocities and spin-dependent scattering off sample boundaries.

In Section 9.2 we introduce the model Hamiltonian of the HgTe/CdTe QW and describe the formalism necessary to calculate the Seebeck and spin Nernst coefficients, and also discuss a generalization of the Mott relation. The thermoelectric transport by the edge states – with a particular emphasis on finite size effects – is analysed in Section 9.3 through the behaviour of Seebeck and spin Nernst signals. In Section 9.4, we focus on the spin-dependent thermoelectric effect induced by the bulk states.

9.2 Model

In this section, we present the model Hamiltonian of the HgTe/CdTe QW and give the general expressions of the Seebeck and spin Nernst coefficients.

9.2.1 Hamiltonian

We consider a four-terminal cross-bar setup based on a HgTe/CdTe QW whose low-energy dynamics is described by the Bernevig-Hughes-Zhang (BHZ) four-band model [Ber+06a; Kön+08]. The Hamiltonian is written in the basis of the lowest QW subbands $|E+\rangle$, $|H+\rangle$, $|E-\rangle$, and $|H-\rangle$. Here, $\pm$ stands for two Kramers partners but in the following, we will refer to them as spin components, denoted by $\uparrow$,$\downarrow$, for brevity. The spin z-direction corresponds to the QW

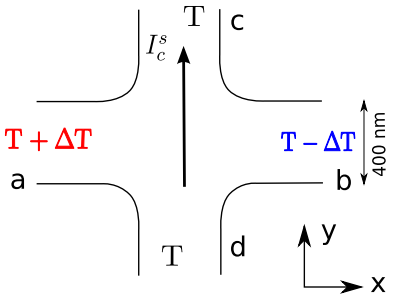


Figure 9.1: Four-terminal cross-bar setup based on HgTe/CdTe QWs used for thermoelectric transport. A longitudinal temperature gradient ΔT is applied between reservoirs a and b and generates a transverse spin current I_c^s detected, for instance, in reservoir c.

growth direction, which is [001]. The Hamiltonian can be written as

$$H = V_m(\mathbf{r})\tau_z - Dk^2 + \begin{pmatrix} h(k) & 0 \\ 0 & h^*(-k) \end{pmatrix}, \tag{9.1}$$

$$h(k) = \begin{pmatrix} \mathcal{M}(k) & Ak_+ \\ Ak_- & -\mathcal{M}(k) \end{pmatrix} \tag{9.2}$$

with $k^2 = k_x^2 + k_y^2$, $k_\pm = k_x \pm ik_y$, and $\mathcal{M}(k) = M - Bk^2$. The sign of the gap parameter M determines whether we are in the regime of a trivial insulator ($M > 0$) or a topological insulator ($M < 0$). Experimentally, M is tuned by changing the QW width.

The term $V_m(\mathbf{r})\tau_z$ describes an in-plane confinement potential, where τ_z is a Pauli matrix acting on the E/H space. By this kind of confinement we may ensure that outside of the sample, i.e. in vacuum, the parameter regime is topologically trivial, so that edge states, if present, will be confined. Calling the inside of the sample G, the limit $V_m(\mathbf{r}) \to \infty \ \forall \mathbf{r} \in \partial G$ can be used to make all components of ψ vanish at the sample boundary, the envelope function ψ being the solution of the Dirac equation based on the Hamiltonian (9.1).

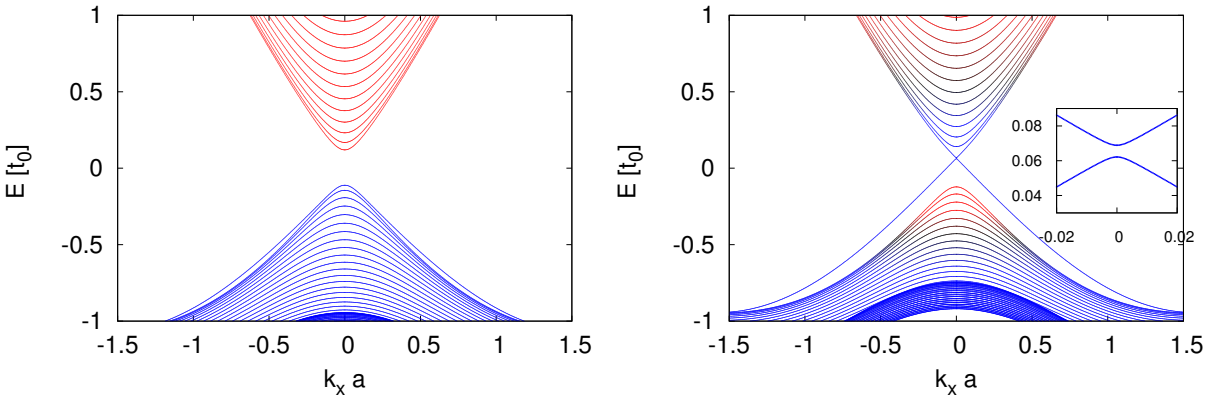
This model can be extended by a term breaking the structural inversion asymmetry (SIA) with a z-dependent potential, as has been shown in Chapter 7.1.1. The resulting Rashba-like interaction connects the Kramers blocks of the Hamiltonian (9.1) affecting the particles with opposite spin

$$h_R(k) = \begin{pmatrix} -iR_0k_- & -iS_0k_-^2 \\ iS_0k_-^2 & iT_0k_-^3 \end{pmatrix} \tag{9.3}$$

with the Rashba coupling parameters R_0, S_0, and T_0 [Rot+10]. We have analysed that such a term will only quantitatively affect all our results presented below. Therefore, we will not further consider effects due to SIA in this chapter.

Figure 9.1 shows the four-terminal cross-bar setup we analyse. The central sample is connected to four semi-infinite leads: the reservoirs a and b are maintained respectively at warmer and colder temperature than the rest of the system – creating a longitudinal temperature gradient – while the transverse terminals c and d are used to probe spin currents.

To model the setup and treat the thermoelectric transport properties, we employ the tight binding approach. Therefore, we discretize the continuum model (9.1) on a lattice of spacing



(a) (b)

Figure 9.2: Subband dispersion relation for the leads (400 nm wide), in (a) the normal regime ($M = 0.1t_0$) and (b) the inverted regime ($M = -0.1t_0$). The inset shows the mini-gap in the dispersion caused by the overlap of edge states. The colouring highlights the transition from electron-like character ($|E\pm\rangle$ in red) to heavy-hole character ($|H\pm\rangle$ in blue) with full color for a 20% excess of either contribution. The lattice constant is $a = 6.6$ nm so $t_0 = 44$ meV. The mini-gap width is $6.5 \times 10^{-3}\ t_0 (= 0.28$ meV). Taken from [Rot+12]. © (2012) by the American Physical Society

a by the substitutions $k_i^2 \to \frac{1}{a^2}(2 - 2\cos k_i a)$ and $k_i \to \frac{1}{a}\sin(k_i a)$, where $i = x, y, z$. The confinement potential is implemented by the lattice truncation in accordance with the geometry of the sample. Rewriting the trigonometric functions in terms of translation operators, this leads to a tight binding Hamiltonian which only contains nearest-neighbour hopping terms between the lattice sites. The energies of the model are expressed as functions of the conduction band hopping parameter $t_0 = -\frac{D+B}{2a^2}$ where the parameter values of the HgTe/CdTe QW are taken as in typical experiments, *i.e.* $A = 0.375$ eVnm, $B = -1.120$ eVnm^2, and $D = -0.730$ eVnm^2. In the low-energy regime, the lattice constant is set sufficiently small compared to the Fermi wave length. Hence, for $a = 6.6$ nm, the energy unit is $t_0 = 44$ meV. The parameter M is chosen as $|M| = 0.1t_0 = 4.4$ meV. In Fig. 9.2(a) and (b), we show the subband dispersion relation for a HgTe/CdTe QW of width 400 nm, both in the normal insulator and the topological insulator regimes. In the latter case, finite size effects emerge on the edge states since they substantially overlap [Zho+08]. One of the consequences of this is the opening of a mini-gap, as shown in the inset of Fig. 9.2(b).

9.2.2 Landauer Büttiker formalism and thermoelectric coefficients

The particle current $I_{p\sigma}$ in the lead p with spin σ is obtained by the Landauer-Büttiker formula [Dat07], which is the generalization of (5.51) to spin-selective leads,

$$I_{p\sigma} = \frac{1}{h}\sum_{q\neq p}\int dE\, T_{p\sigma,q}(E)(f_p - f_q) \tag{9.4}$$

with $f_p = (e^{(E-\mu_p)/k_B \mathrm{T}_p} + 1)^{-1}$ the electronic Fermi distribution function, k_B the Boltzmann constant, T_p the temperature, and μ_p the chemical potential. The transmission probability

$T_{p\sigma,q}(E) = \sum_{\sigma'} T_{p\sigma,q\sigma'}(E)$ from lead p with spin σ to lead q can be evaluated using the non-equilibrium Green's function formalism (NEGF) [Mei+92; San+99; Wim08]

$$T_{p\sigma,q\sigma'}(E) = \mathrm{Tr}[\Gamma_{p\sigma} G^R \Gamma_{q\sigma'} G^A], \quad (p,\sigma) \neq (q,\sigma'), \tag{9.5}$$

where $\Gamma_{p\sigma}(E) = i(\Sigma_{p\sigma} - \Sigma_{p\sigma}^\dagger)$ refers to projectors on velocity operators of the propagating modes, and $\Sigma_{p\sigma}(E)$ stands for the spin-dependent self-energy. Equation (9.5) can also be obtained from the Fisher-Lee relation (5.74). The self-energy is defined by $\Sigma_{p\sigma}(E) = \tau_{p\sigma}(E + i0^+ - H_{\mathrm{leads}})^{-1}\tau_{p\sigma}^\dagger$, where the matrix $\tau_{p\sigma}$ connects the lead p, spin σ to the adjacent sites of the sample. $G^R(E) = (G^A(E))^\dagger = (E - H_{\mathrm{sample}} - \sum_{p,\sigma}\Sigma_{p\sigma})^{-1}$ is the retarded Green's function. Further, H_{leads} and H_{sample} represent, respectively, the lattice Hamiltonians of the decoupled leads and the sample. Once the transmission probabilities are evaluated, the charge current $I_p^e = e(I_{p\uparrow} + I_{p\downarrow})$ and the spin current $I_p^s = (\hbar/2)(I_{p\uparrow} - I_{p\downarrow})$ can be obtained.

We consider a longitudinal temperature gradient between the leads a and b by setting $\mathrm{T}_\mathsf{a} = \mathrm{T} + \Delta\mathrm{T}$, $\mathrm{T}_\mathsf{b} = \mathrm{T} - \Delta\mathrm{T}$, and $\mathrm{T}_\mathsf{c} = \mathrm{T}_\mathsf{d} = \mathrm{T}$. In the linear response regime, the Seebeck coefficient reports the longitudinal voltage bias $\Delta\mu = \frac{\mu_\mathsf{a} - \mu_\mathsf{b}}{2}$ generated by the temperature gradient $\Delta\mathrm{T}$ under the condition of vanishing charge currents (open boundary conditions). Upon Taylor expansion of the Fermi functions in $\Delta\mathrm{T}$ and $\Delta\mu$, the Seebeck coefficient can be written as

$$S_e = \left.\frac{-\Delta\mu}{e\Delta\mathrm{T}}\right|_{I^e_{\mathsf{a},\mathsf{b}}=0} \approx \frac{1}{e\mathrm{T}} \frac{\int dE\, f_0(1-f_0)\mathcal{T}_{SE}(E)(E-\mu)}{\int dE\, f_0(1-f_0)\mathcal{T}_{SE}(E)} \tag{9.6}$$

with $f_0 = (e^{(E-\mu)/k_B\mathrm{T}} + 1)^{-1}$ the Fermi distribution function at equilibrium. In the above equation, we defined the Seebeck transmission function $\mathcal{T}_{SE}(E) = T_{\mathsf{a},\mathsf{b}} + (T_{\mathsf{a},\mathsf{c}} + T_{\mathsf{a},\mathsf{d}})/2$, where the summation over spins is implied.

Due to the presence of intrinsic spin-orbit interaction in the sample, the longitudinal thermal gradient $\Delta\mathrm{T}$ also yields a transverse spin current $I^s_\mathsf{c} (= -I^s_\mathsf{d})$ in the case of closed boundary conditions. The spin Nernst coefficient is then defined as the ratio

$$N_s = \left.\frac{I_c^s}{2\Delta\mathrm{T}}\right|_{\mu_{\mathsf{c},\mathsf{d}}=\mu} \approx \frac{1}{8\pi k_B \mathrm{T}^2} \int dE\, f_0(1-f_0)\mathcal{T}_{SN}(E)(E-\mu). \tag{9.7}$$

Here, we introduced the spin Nernst transmission function $\mathcal{T}_{SN}(E) = \Delta T_{\mathsf{c},\mathsf{b}} - \Delta T_{\mathsf{c},\mathsf{a}}$, with the short-hand notation $\Delta T_{\mathsf{c},\mathsf{b}}(E) = T_{\mathsf{c}\uparrow,\mathsf{b}\uparrow} + T_{\mathsf{c}\uparrow,\mathsf{b}\downarrow} - T_{\mathsf{c}\downarrow,\mathsf{b}\uparrow} - T_{\mathsf{c}\downarrow,\mathsf{b}\downarrow}$.

Interestingly, the Mott relation provides information about the (spin-)thermotransport coefficients on the basis of the energy dependence of the (spin-)conductance [Cut+69]. In the low temperature limit, one can derive

$$S_e \approx \frac{\pi^2 k_B^2 \mathrm{T}}{3e} \left.\frac{d\ln G_{xx}(E)}{dE}\right|_{E=\mu}, \tag{9.8}$$

where $G_{xx}(E, \mathrm{T}=0) = \frac{e^2}{h}\mathcal{T}_{SE}(E)$ is the longitudinal conductance for zero temperature. Equation (9.8) is valid if $k_B\mathrm{T}$ is large compared to the scale on which $\mathcal{T}_{SE}(E)$ varies. A numerical analysis in [Lun+05] claims that (9.8) can be valid even if $\mathcal{T}_{SE}(E)$ varies more rapidly, as long as $k_B\mathrm{T} \ll \mu$.

An analogous relation exists between the spin Nernst signal and the spin Hall conductance. From the Sommerfeld expansion of the transmission function in Eq. (9.7), one obtains the

following Mott-like formula

$$N_s \approx \frac{2\pi^2 k_B^2 \mathrm{T}}{3e} \left.\frac{dG_{sH}(E)}{dE}\right|_{E=\mu(\mathrm{T}=0)} \tag{9.9}$$

with $G_{sH}(E, \mathrm{T}=0) = \frac{e}{8\pi}\mathcal{T}_{SN}(E)$, the spin Hall conductance at zero temperature. In the next subsection, we demonstrate that this relation can be extended to finite temperature by defining a smoothed function $\tilde{\mathcal{T}}_{SN}(E)$ (see Eq. (9.14)) that depends on the temperatures in the leads.

9.2.3 Mott-like relation

In this section, we show how the Mott-like relation of Eq. (9.9) can be generalized to finite temperatures. For this we consider the Fourier transform of transmission functions $\Delta T_{\mathsf{c},q}(\tau) = \frac{1}{2\pi}\int dE e^{-iE\tau}\Delta T_{\mathsf{c},q}$. The spin Nernst effect is defined as

$$N_s = \frac{I^s_{\mathsf{c}}}{2\Delta \mathrm{T}} = \frac{1}{8\pi\Delta \mathrm{T}}\int dE \sum_q \Delta T_{\mathsf{c},q}(E)(f_{\mathsf{c}} - f_q), \tag{9.10}$$

where the potential μ is assumed to be the same for all leads, while the temperatures may differ. The integral has the form of a convolution, since f_{c} and f_q are functions of $E-\mu$. Therefore, the integrand of the Fourier representation is a product of functions of τ,

$$N_s = \frac{1}{4\pi}\int_{-\infty}^{\infty} d\tau \frac{i}{\tau} e^{-i\mu\tau} \sum_q \frac{\Delta T_{\mathsf{c},q}(\tau)}{\Delta \mathrm{T}} \left.\frac{x}{\sinh x}\right|_{x=\pi\tau k_B \mathrm{T}_q}^{x=\pi\tau k_B \mathrm{T}_{\mathsf{c}}}. \tag{9.11}$$

Further, we define a symmetric "smoothing" function that depends on temperatures of the leads c and q as

$$F^q(\tau) = \frac{3}{\pi^2 k_B^2 \mathrm{T}_{\mathsf{c}}\Delta \mathrm{T}} \frac{1}{\tau^2} \left.\frac{x}{\sinh x}\right|_{x=\pi\tau k_B \mathrm{T}_q}^{x=\pi\tau k_B \mathrm{T}_{\mathsf{c}}}. \tag{9.12}$$

If we put $\mathrm{T}_{\mathsf{c}} - \mathrm{T}_q = \Delta \mathrm{T}$, we find with $\Delta \mathrm{T} \to 0$

$$F^q(\tau) \approx 3 \left.\frac{x\coth x - 1}{x\sinh x}\right|_{x=\pi\tau k_B \mathrm{T}_{\mathsf{c}}}, \tag{9.13}$$

which has a width of $\Delta\tau \approx \frac{4}{\pi k_B \mathrm{T}_{\mathsf{c}}}$ and $F^q(0) \to 1$. Now, we define a temperature-smoothed spin Nernst transmission function as

$$\tilde{\mathcal{T}}_{SN}(\tau) = \sum_q F^q(\tau)\Delta T_{\mathsf{c},q}(\tau), \tag{9.14}$$

which implies that $\tilde{\mathcal{T}}_{SN}(E) = \int d\tau e^{-iE\tau}\tilde{\mathcal{T}}_{SN}(\tau)$ is real, and compared to the original function $\mathcal{T}_{SN}(E)$, it is smoothed on the scale of $k_B \mathrm{T}_{\mathsf{c}}$. Finally, we obtain the relation

$$N_s(\mu) = \frac{\pi k_B^2 \mathrm{T}_{\mathsf{c}}}{12} \left.\frac{d\tilde{\mathcal{T}}_{SN}(E)}{dE}\right|_{E=\mu}, \tag{9.15}$$

which is exact to all orders in T_{c} and $\Delta \mathrm{T}$. The meaning of the latter equation is the following one: First taking the derivative $\partial_E \mathcal{T}_{SN}$ and then smoothing by temperature is the same as first smoothing with a modified smoothing kernel and then taking the derivative. Since $\mathcal{T}_{SN}(E)$ shows a highly oscillating behaviour, the above equation simplifies the interpretation of the spin Nernst signal in terms of transmission functions because of the smoothing of $\tilde{\mathcal{T}}_{SN}(E)$. Equation (9.15) is one of the key results of this chapter.

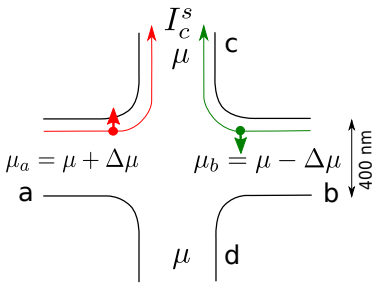


Figure 9.3: Four-terminal cross-bar setup based on a HgTe/CdTe QW in the inverted regime. When an electrical bias is imposed between the longitudinal leads a and b, one edge channel carries electrons with spin up from lead a to lead c (red solid line) and one edge channel carries electrons with spin down from lead b to lead c (green solid line). This gives rise to a spin current in lead c and, at zero temperature, to a quantized spin Hall conductance. Taken from [Rot+12]. © (2012) by the American Physical Society

9.3 Thermoelectric transport carried by the edge states

In this section, we present the numerical results of the spin Nernst and Seebeck coefficients for an energy regime within the bulk gap. When the HgTe/CdTe QW is in a topologically trivial phase, there is no sub-gap transport through the system. The transmission functions $\mathcal{T}_{SE}$ and $\mathcal{T}_{SN}$ are then zero, and from Eqs. (9.6) and (9.7), it follows that there are no thermoelectric signals. On the contrary, the HgTe/CdTe QW in a non trivial phase hosts edge states in the bulk insulating gap. These modes carry electrons with opposite spins in opposite directions.

With respect to the geometry of the setup, spin and electrical currents are induced and flow respectively in transverse and longitudinal leads, as depicted in Fig. 9.3. However, the finite width of the system implies an overlap of the edge states meaning that backscattering processes can occur.

We first investigate the behaviour of the spin Nernst signal N_s and the associated transmission function $\mathcal{T}_{SN}$. The results are presented in Figs. 9.4(a) and (b). While the chemical potential is in the bulk gap, the spin transport is mediated by the edge channels so that the transmission function is simply given by

$$\mathcal{T}_{SN} = (T_{\mathsf{c}\uparrow,\mathsf{b}} - T_{\mathsf{c}\downarrow,\mathsf{b}}) - (T_{\mathsf{c}\uparrow,\mathsf{a}} - T_{\mathsf{c}\downarrow,\mathsf{a}}) = -2. \tag{9.16}$$

Evidently, as the chemical potential reaches the boundary of the mini-gap, the number of propagating states drops to zero and transport breaks down. This results in a peak of the transmission function $\mathcal{T}_{SN}$. Consequently, the spin Nernst coefficient is zero in the bulk gap except when the chemical potential reaches the boundary of the mini-gap. Because of the Mott-like relation (9.15), a symmetric function $\mathcal{T}_{SN}(E)$ must result in an antisymmetric function $N_s(E)$. Therefore, N_s exhibits an approximately antisymmetric peak centred at the maximum of the transmission peak. The confinement of the QW implies an energy shift in the band dispersion. Therefore, the boundaries of the bulk gap are not exactly at energy $|M| = 0.1t_0$, as we can see in Figs. 9.2 and 9.4. In Fig. 9.4, the gap and minigap positions of a finite system are indicated by vertical lines. Dotted vertical lines are used for the inverted regime and dashed lines for the normal regime. Interestingly, one observes that the merging of the edge state to the conduction band causes $\mathcal{T}_{SN}$ to vanish already before the first bulk mode appears. Where μ lies between

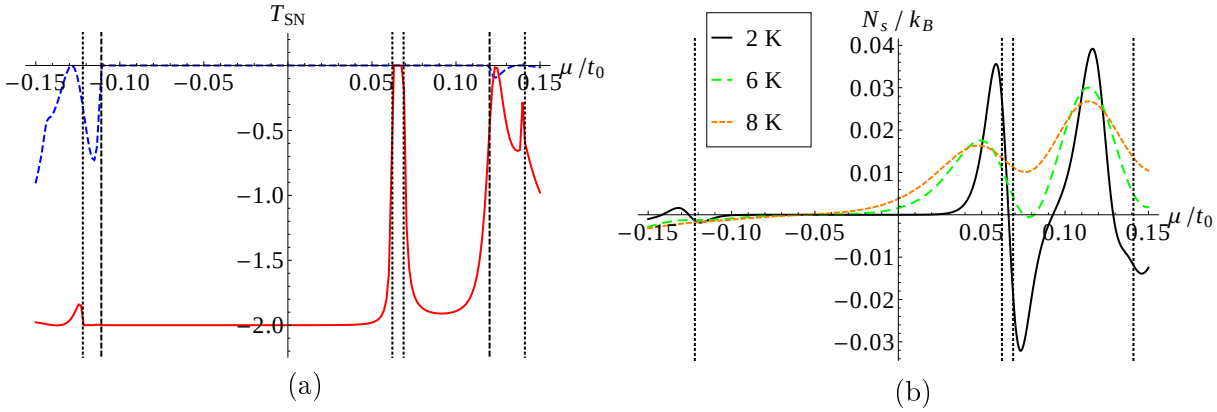


Figure 9.4: (a) Spin Nernst transmission function as a function of the chemical potential when the sample is in the normal (dashed blue line) or in the inverted (solid red line) regime. The dotted vertical lines indicate the bulk gap and minigap positions in a finite system for the inverted regime, while the dashed vertical lines indicate the gap in the normal regime. (b) Spin Nernst signal N_s/k_B in a system in the inverted regime at T = 2 K (black solid line), T = 6 K (green dashed line), and T = 8 K (orange dotted line). Taken from [Rot+12]. © (2012) by the American Physical Society

the right dashed and dotted vertical lines, a finite $\mathcal{T}_{SN}$ reappears due to the formation of the first bulk state, in the same subband as the edge state. Outside the gap indicated by the vertical lines, bulk states start to participate to the spin transport resulting in additional oscillations in $\mathcal{T}_{SN}$ as a function of μ. They transform into peaks of the spin Nernst coefficient whose existence is understood with the same arguments as for the mini-gap peak. Especially at positive chemical potential, the magnitude of the peak is comparable to that of the mini-gap and allows to mark the position where the edge states merge.

In Fig. 9.4(b), we show the behaviour of the spin Nernst coefficient for different temperatures. As k_BT increases, the position of the peaks is slightly shifted to lower energy. The magnitude tends to decrease and the peak width is broadened. Up to T = 6 K, the spin Nernst signal goes to zero between the peak that specify the position of the mini-gap and the edge state merging peak. Beyond this temperature, N_s is smoothed out, so that it can not probe the edge state signal.

We now turn to the analysis of the transmission function $\mathcal{T}_{SE}$ and the Seebeck coefficient S_e as a function of energy. The results are presented in Fig. 9.5. Inside the bulk gap, the transmission function $\mathcal{T}_{SE}$ is constant but goes to zero when the chemical potential is in the mini-gap. This feature leads to an approximately antisymmetric peak in the behaviour of S_e, which provides information on the presence and the position of the mini-gap in the spectrum. The boundary of the bulk gap manifests itself as the step of the transmission function and transforms as a narrow peak in S_e.

The transmission function $\mathcal{T}_{SE}$ exhibits a smoothed staircase behaviour whose steps coincide with the opening (at positive energy) or the closing (at negative energy) of conducting channels. This behaviour transforms into a series of peaks in S_e. However, as the chemical potential increases, the magnitude of the peaks reduces. The reason is that the considered setup possesses four terminals that all exhibit an increasing number of propagating modes with increasing μ. Thus, inter-mode scattering is more and more likely to happen. Then, the staircase behaviour

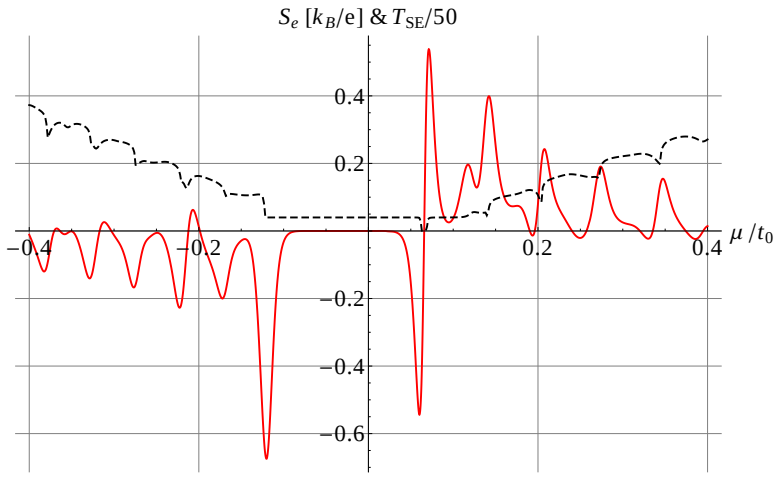


Figure 9.5: Seebeck coefficient $S_e[k_B/e]$ (red solid line) and scaled transmission function $\mathcal{T}_{SE}/50$ (black dashed line) as a function of the chemical potential at T = 2 K. The mini-gap appears as an antisymmetric peak. Taken from [Rot+12]. © (2012) by the American Physical Society

of $\mathcal{T}_{SE}$ diminishes and transforms into oscillations.

We close this section with a remark on the average sign of the Seebeck coefficient S_e. It is positive in the conduction band and negative in the valence band which reflects the sign of the corresponding charge excitations in a given band.

9.4 Spin Nernst effect induced by bulk states

A spatial dependence of model parameters, like, for instance, an in-plane electrostatic potential V_0 or the mass confinement potential $\tau_z V_m$, can generate a transverse spin current resulting in a spin Hall signal for the metallic bulk states (see Chapter 7). This phenomenon has been previously analysed in Refs. [Yok+09; Gui+11; Yam+11] in the context of charge and spin transport properties at interfaces between metals and quantum spin Hall systems. As already mentioned above, the spin Hall conductance gives rise to the spin Nernst signal from the bulk states through the Mott-like relation (9.9). Therefore, in the next two subsections we will focus on analytical models to describe the scaling of the spin current and spin Hall conductance with the band structure parameters and compare our intuitive analytical models with the numerics.

First however, to visualize the formation of the spin Hall effect at the sample boundary, it is instructive to plot the local spin current density in the numerical four-band model. In order to do so, we first define a local spin current operator by

$$\hat{\mathbf{J}}^z(\mathbf{r}') = \frac{1}{i}\{[\hat{\mathbf{r}}, H], \delta(\hat{\mathbf{r}} - \mathbf{r}')\}\sigma_z, \tag{9.17}$$

where σ_z is a Pauli matrix that acts on the spin space of the four-band model. On the basis of the NEGF, it is then straightforward to evaluate the expectation value of the spin current operator at T = 0, which can be expressed as

$$\mathbf{J}^z(\mathbf{r}) = \sum_p \mathrm{Tr}\left[\hat{\mathbf{J}}^z(\mathbf{r}) G^R \Gamma_p G^A\right] \mu_p. \tag{9.18}$$

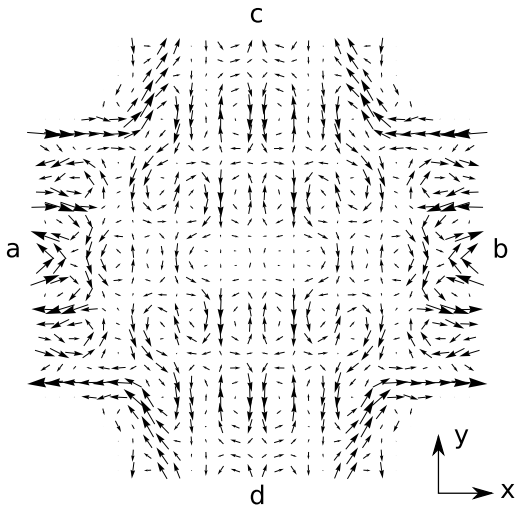


Figure 9.6: Local spin current $\mathbf{J}^z$ for normal metallic regime, and electrical bias, $\mu_\mathsf{a} - \mu = -(\mu_\mathsf{b} - \mu)$ with $\mu = 0.4t_0$ and T = 0. Taken from [Rot+12]. © (2012) by the American Physical Society

In Fig. 9.6, the local spin current is shown for the normal metallic bulk regime and an electrical bias applied from left to right. One clearly recognizes a spin current flowing along the edges.

Note that our model shows local spin currents already at equilibrium. However at equilibrium, the spin current integrated over the cross section of a lead cancels and thus does not enter the spin Hall signal. For clarity, the local spin current that we show in Fig. 9.6 is only the non-equilibrium part.

The rest of this section is organized as follows: 9.4.1 and 9.4.2 deliver two complementary approaches to explain the interface spin current that is transverse to the potential gradient. In part 9.4.3, the connection between these interface spin currents and the spin-Nernst coefficient gives us a qualitative understanding of the behaviour of spin-thermo effects for the bulk states.

9.4.1 Effective two-band models

Here, we show that an anomalous spin-dependent velocity naturally appears within effective two-band spin-diagonal electron or hole band models obtained by perturbatively folding down the four-band model (see Eq. (9.1)). We apply third order quasi-degenerate perturbation theory similar to our previous work of Chapter 7.2. The diagonal part of the Hamiltonian is $h_0 = \mathrm{diag}(M, -M)$. For the perturbation part, we consider $h'^{\uparrow} = \begin{pmatrix} \tilde{V}_e & A\hat{k}_+ \\ A\hat{k}_- & \tilde{V}_h \end{pmatrix}$ and $h'^{\downarrow} = h'^{\uparrow *}(-k)$, where $\tilde{V}_{e/h} = V_{e/h} \mp Bk^2 - Dk^2$ and $V_{e/h} = V_0 \pm V_m$; the subscripts e and h refer to electron and heavy hole bands, respectively. Note that, as compared to Eq. (9.1), we allow for a finite in-plane potential $V_0(x, y) = V_0(\mathbf{r})$ in this analysis. $V_0(\mathbf{r})$ models an electrostatic potential that acts equally on electron and hole bands, while $V_m(\mathbf{r})$ models a mass-like confinement acting with opposite signs on electron and hole bands. The B and D parameters will not enter to the spin current in third order perturbation theory. Treating $\hat{k}_\pm$ as operators acting on a perturbing potential, we obtain the spin-dependent effective two-band Hamiltonians as follows (showing

terms up to third order in perturbation theory)

$$h^{\uparrow}_{\mathrm{eff,e}} = V_e - (D+B)\hat{k}^2 + \frac{1}{8M^2}\left(2A^2\hat{k}_+\tilde{V}_h\hat{k}_- - \{A^2\hat{k}^2, \tilde{V}_e\}\right), \tag{9.19}$$

$$h^{\uparrow}_{\mathrm{eff,h}} = V_h + (D-B)\hat{k}^2 + \frac{1}{8M^2}\left(2A^2\hat{k}_-\tilde{V}_e\hat{k}_+ - \{A^2\hat{k}^2, \tilde{V}_h\}\right), \tag{9.20}$$

and $h^{\downarrow}_{\mathrm{eff,e/h}} = (h^{\uparrow}_{\mathrm{eff,e/h}})^*$. The lowest order spin-dependent term of the effective electron/heavy hole model is thus given by

$$h_{\mathrm{Pauli,e/h}} = \frac{1}{2}\left(h^{\uparrow}_{\mathrm{eff,e/h}} - h^{\downarrow}_{\mathrm{eff,e/h}}\right)\sigma_z = -\frac{A^2}{4M^2}\left(\nabla(\pm V_0 - V_m)\times\mathbf{k}\right)_z\sigma_z. \tag{9.21}$$

In the Heisenberg picture, we obtain a spin-dependent anomalous velocity

$$\mathbf{v}'^{\mathrm{an,e/h}} = \frac{1}{i\hbar}[\mathbf{r}, h_{\mathrm{Pauli,e/h}}] = \frac{A^2}{4M^2}\sigma_z\begin{pmatrix} -\partial_y \\ \partial_x \\ 0 \end{pmatrix}(\pm V_0 - V_m). \tag{9.22}$$

Since $\mathbf{v}'^{\mathrm{an,e/h}} \perp \nabla(V_0 \mp V_m)$, we expect to see a spin current along the edge of the sample, similarly to the spin current carried by the edge states, but now the effect is induced by the bulk modes.

Although this gives us an idea of the mechanism that generates the spin current, $\mathbf{v}'^{\mathrm{an}}$ cannot be used for any quantitative comparison with numerical results that we obtain for the four-band model, since neither V_0 nor V_m are included in the numerical code. We introduced them to model the electrostatic confinement, which is simply ensured by the lattice truncation in the numerical model. As long as $E_F \neq 0$, we can expect that the real physical situation corresponds to a mixture of potentials V_0 and V_m. V_0 alone is not suitable for confinement of carriers, but it models an electrostatic field that builds up at the sample edge. On the other hand, at the sample edge the electronic gap $2M$ must increase until the vacuum gap $2m_0c^2$ is reached. This will cause efficient confinement of carriers (and even edge states) and can be modelled by V_m. Although the potentials V_0 and V_m must be large for ensuring confinement, they are treated as perturbation (small compared to $2M$) in (9.22).

With regard to quantitative results, there is another issue with (9.22). We already analysed the anomalous velocity generated by an electrostatic potential V_0 in Section 4.2. There, we concluded that for consistent results, the non-commuting space operator found by projection of the Dirac space operator on the electron band should be used, $\hat{\mathbf{r}}_c = \mathbf{r} + \mathbf{A}^+$, which includes the Berry connection for the electron band given by $\mathbf{A}^+ = \frac{A^2}{2E(k)(E(k)+M)}\mathbf{k}\times\boldsymbol{\sigma}$. Here we replaced the parameters of the Dirac model by those of the four-band model, with dispersion $E(k)$. We interpreted the non-commuting Berry connection term as fingerprint of the relativistic transformation of a spinning particle. Without the non-commuting part, the operator $\mathbf{r}$ that appears in the Pauli equation should be understood as mean coordinate, and the resulting anomalous velocity will be a different physical quantity. Literature seems to agree that the Berry correction in $\hat{\mathbf{r}}_c$ should be included for rigorous calculations e.g. of the spin Hall and anomalous Hall effects [Chu+10; Xia+10; Han+09b]. Repeating the analysis of Section 4.2 to find the part of the anomalous velocity due to $\tau_z V_m = \tau_z\mathbf{r}\cdot\nabla V_m$, we find that the projection on the positive-energy band is given by $PU(k)\tau_z\mathbf{r}U^\dagger(k)P = \frac{M}{E(k)}\mathbf{r} - \mathbf{A}^+$, compared to $\hat{\mathbf{r}}_c = PU(k)\mathbf{r}U^\dagger(k)P = \mathbf{r} + \mathbf{A}^+$, where $U(k)$ is the unitary rotation that diagonalizes the four-band

model. The expression for the anomalous velocity involves a term similar to the generalized Berry curvature (4.48),

$$\frac{1}{i\hbar}[\hat{\mathbf{r}}_c, PU\tau_z V_m U^\dagger P] = \sum_l \frac{-1}{\hbar}\left(\partial_{k_j} A_l^+ + \frac{M}{E(k)}\partial_{k_l} A_j^+ + \frac{1}{i}[A_j^+, A_l^+]\right)\partial_l V_m \tag{9.23}$$

$$= \sum_l \frac{1}{\hbar}\left(\left(\frac{M}{E(k)} - 1\right)\partial_{k_j} A_l^+ - \frac{1}{i}[A_j^+, A_l^+]\right)\partial_l V_m. \tag{9.24}$$

The commutator term does not have the form of the anomalous velocity, but evaluates to $\frac{1}{i}[A_j^+, A_l^+] = \frac{A^4}{2E^2(k)(E(k)+M)^2}\epsilon_{jlm}k_m(\mathbf{k}\cdot\boldsymbol{\sigma})$. Approximating $1 - \frac{M}{E(k)} \approx \frac{E_{\text{kin}}}{M}$, we find

$$\mathbf{v}^{\text{an,e}} = \frac{A^2}{4M^2}\sigma_z \begin{pmatrix} -\partial_y \\ \partial_x \\ 0 \end{pmatrix}\left(2V_0 - \frac{E_{\text{kin}}}{M}V_m\right). \tag{9.25}$$

Compared to (9.22), the contribution due to ∇V_0 is enhanced by a factor of two, while the contribution due to ∇V_m is suppressed by the factor $\frac{E_{\text{kin}}}{M}$.

Note that the assumptions for a valid perturbation theory are quite restrictive. The condition $Ak \ll 2|M|$ restricts the energy range to about $|E| \lesssim 0.2t_0$. Further, this approach works in the inverted regime only when one considers the bulk states and assumes a direct gap. The main drawback of this perturbative approach is, however, that it assumes the variation of the potentials V_0, V_m small compared to the gap $2|M|$, which is not the case for the numerical confinement potential. Therefore, although we expect to find qualitative results by this approach, it is important to compare it with the non-perturbative model including hard wall boundary conditions which will be done in the next subsection.

9.4.2 Hard wall boundary spin current

In this subsection, we present a complementary explanation of the spin current carried by the bulk states, valid also beyond the parameter regime $Ak \ll 2|M|$, demonstrating that the reflection of an incident wave at a hard wall boundary leads to a spin current along the boundary. Due to a phase offset, this spin current persists even for a superposition of waves incident at different angles. We will show below that in the regime $Ak \ll 2|M|$ the spin current scales like A^2/M. Interestingly, we observe that the explanation of the spin current given here seems to be close to what is seen in the numerical four-band tight binding model, because the numerically calculated spin Hall effect indeed scales like A^2/M in the parameter regime $Ak \ll 2M$ (with $M > 0$).

In the model we consider now, the hard wall boundary condition for the envelope function is given by $\psi(y = 0) = 0$. While the direction of the outgoing beam is restricted by the energy and momentum conservation laws and is not spin-dependent, there is a spin-dependent phase shift between incident and reflected wave. Remarkably, even for an incident wave normal to the interface, a spin current moving along the interface is generated. In case this interface is bent, like it happens at the sample boundary connecting two perpendicular leads, it will transform into a spin Hall signal (like in Fig. (9.6)).

We start with the following ansatz for the spin $\uparrow$ wave function

$$\psi^\uparrow(y) = \langle y|\psi^\uparrow\rangle = e^{ik_y y}u_{k_y} + re^{-ik_y y}u_{-k_y} + ce^{\lambda y}u_{i\lambda}, \tag{9.26}$$

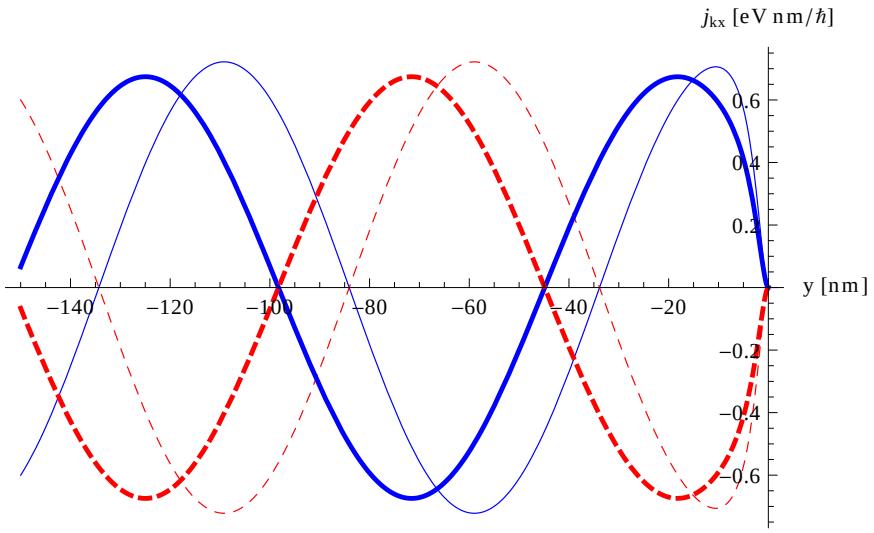


Figure 9.7: The spin-dependent currents parallel to the interface at $y = 0$ are plotted, with $j^{\uparrow}_{k_x}(y)$ in blue (solid) and $j^{\downarrow}_{k_x}(y)$ in red (dashed). For perpendicular angle of incidence ($k_x = 0$), the phase difference $\Delta\phi$ is always π. The energy is $E = 0.3t_0$ in all cases. The thick/thin lines show the normal/inverted regime with $M = 0.1t_0$ and $M = -0.1t_0$, respectively. The other parameters are the same as for the lattice model. Taken from [Rot+12]. © (2012) by the American Physical Society

where the plane wave dependency on x has been separated off. $u_{\pm k_y}$ and $u_{i\lambda}$ denote the spinors for fixed energy E and momentum k_x. The condition $\psi^{\uparrow}(0) = 0$ gives two equations for the coefficients r and c. The corresponding coefficients for spin down can be found by replacing $k_x \to -k_x$. The operators of transverse velocity, $V_x^{\uparrow}(k_x) = \frac{1}{\hbar}\frac{\partial h^{\uparrow}}{\partial k_x}$ and $V_x^{\downarrow}(k_x) = -V_x^{\uparrow}(-k_x)$ are independent of k_y and complex-valued matrices. In the following, we will plot both spin up (in blue) and spin down currents (in red), evaluated by

$$j^{\uparrow}_{k_x}(y) = \langle\psi^{\uparrow}|\delta(y-\hat{y})V_x^{\uparrow}|\psi^{\uparrow}\rangle,\ j^{\downarrow}_{k_x}(y) = -j^{\uparrow}_{-k_x}(y). \tag{9.27}$$

It is easy to see that the spin current

$$j^s(y) = j^{\uparrow}_{k_x}(y) - j^{\downarrow}_{k_x}(y) = j^{\uparrow}_{k_x}(y) + j^{\uparrow}_{-k_x}(y) \tag{9.28}$$

is symmetric in the angle of incidence $\theta = \tan^{-1}\frac{k_x}{k_y}$.

Figure 9.7 shows the spin up and down currents for typical parameters and the energy in the conduction band. The superposition of incoming and reflected propagating waves leads to an oscillating pattern. We are interested in the phase shift between spin up and down. The direct terms in $j^{\uparrow}_{k_x}(y)$ (i.e. two incoming or two outgoing propagating modes) are constant in y and current conservation dictates that the incoming and reflected currents are the same. Rotational invariance of the BHZ Hamiltonian and current conservation dictate that $|r^2| = 1$ independent of the spin. Because of time reversal symmetry, the current of the direct terms is independent of the spin and thus, the direct terms do not contribute to $j^s(y)$.

The interference term between the incoming and outgoing modes in $j^{\uparrow}_{k_x}(y)$ is given by

$$2\mathrm{Re}\left[\langle u_{k_y}|V_x^{\uparrow}|u_{-k_y}\rangle r e^{-2ik_y y}\right] = \left|\langle u_{k_y}|V_x^{\uparrow}|u_{-k_y}\rangle r\right| 2\cos(2k_y y - \phi_1^{\uparrow} - \phi_2^{\uparrow}), \tag{9.29}$$

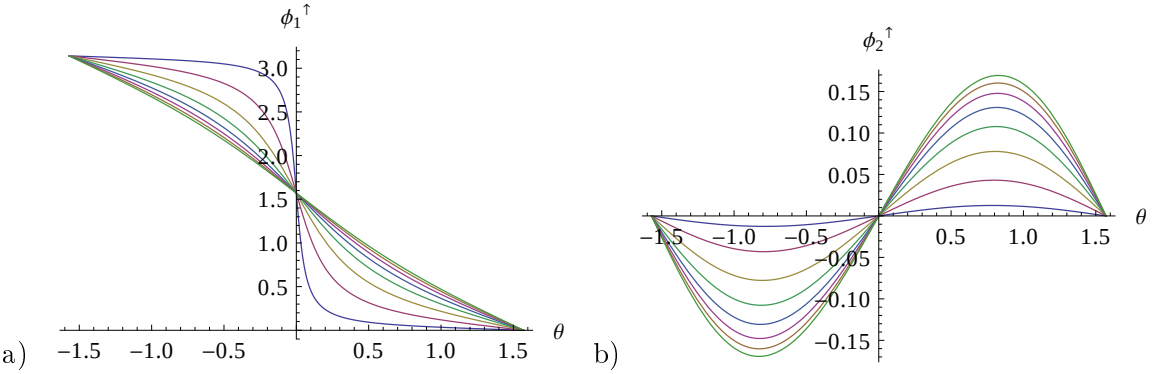


Figure 9.8: The spin up phase shift of the reflected current is $\phi_1^{\uparrow} + \phi_2^{\uparrow}$. a) The phase (of the velocity matrix element) $\phi_1^{\uparrow}$ as a function of the angle of incidence. It has the symmetry $\phi_1^{\uparrow}(-\theta) = \pi - \phi_1^{\uparrow}(\theta)$. Different colors (blue,red,orange, ...) correspond to different choices of $A = 0.05, 0.1, ..., 0.4$eV nm, where the limit $A \to 0$ gives a step function. Parameter values are $E = 0.3t_0$, $M = 0.1t_0$ and B, D have values as in Section 9.2.1. The picture does not change qualitatively, if we change the parameters of the underlying model. b) The phase (of the reflection coefficient) $\phi_2^{\uparrow}$ for the same parameters. For low values of A, it is proportional to A^2, whereas for large values of A it saturates. Taken from [Rot+12]. © (2012) by the American Physical Society

where $\phi_1^{\uparrow} = \arg(\langle u_{k_y}|V_x^{\uparrow}|u_{-k_y}\rangle)$ and $\phi_2^{\uparrow} = \arg(r)$. In Ref. [Yok+09], $\phi_2^{\uparrow} - \phi_2^{\downarrow}$ is called the angle of giant spin rotation. At $k_x = 0$, we have

$$\langle u_{k_y}|V_x^{\uparrow}|u_{-k_y}\rangle\Big|_{k_x=0} = \frac{iA^2k_y}{\sqrt{A^2k_y^2 + (M - Bk_y^2)^2}}, \tag{9.30}$$

where k_y is fixed by the energy. A first-order expansion in k, valid in the regime $Ak \ll 2|M|$ yields

$$j_s(y) \propto \frac{A^2k}{|M|}. \tag{9.31}$$

In contrast to [Gui+11], the spin current in our analytical analysis is connected only with the propagating solutions as explained above. As one can see from Fig. 9.7, where the evanescent modes are included, the periodicity of $j_{k_x}^{\uparrow}(y)$ and $j_{k_x}^{\downarrow}(y)$ is only slightly affected which means that the evanescent contribution at least for the normal regime is minor and Eq. (9.31) still holds.

Let us now analyse the phase relations between spin up and spin down currents more closely. The two phases $\phi_1^{\uparrow}$ and $\phi_2^{\uparrow}$ (defined above) behave differently as a function of the angle of incidence, as shown in Fig. 9.8. We find the symmetries $\phi_1^{\uparrow}(\theta) = \pi - \phi_1^{\uparrow}(-\theta)$ and $\phi_2^{\uparrow}(\theta) = -\phi_2^{\uparrow}(-\theta)$. For $A \to 0$, $\phi^{\uparrow}(\theta) = \phi_1^{\uparrow}(\theta) + \phi_2^{\uparrow}(\theta)$ becomes a step function, with $\phi^{\uparrow}(0) = \pi/2$. We are interested in $\Delta\phi = \phi^{\uparrow} - \phi^{\downarrow}$. For this, we again use a symmetry. If we flip the spin, $r^{\downarrow}(k_x) = r^{\uparrow}(-k_x)$ implies that $\phi_2^{\downarrow}(-\theta) = \phi_2^{\uparrow}(\theta)$ and $V^{\downarrow}(k_x) = -V^{\uparrow}(-k_x)$ implies that $\phi_1^{\downarrow}(-\theta) = \pi + \phi_1^{\uparrow}(\theta)$. Thus,

$$\Delta\phi = \phi_1^{\uparrow} + \phi_2^{\uparrow} - \phi_1^{\downarrow} - \phi_2^{\downarrow} = 2(\phi_1^{\uparrow} + \phi_2^{\uparrow}) \tag{9.32}$$

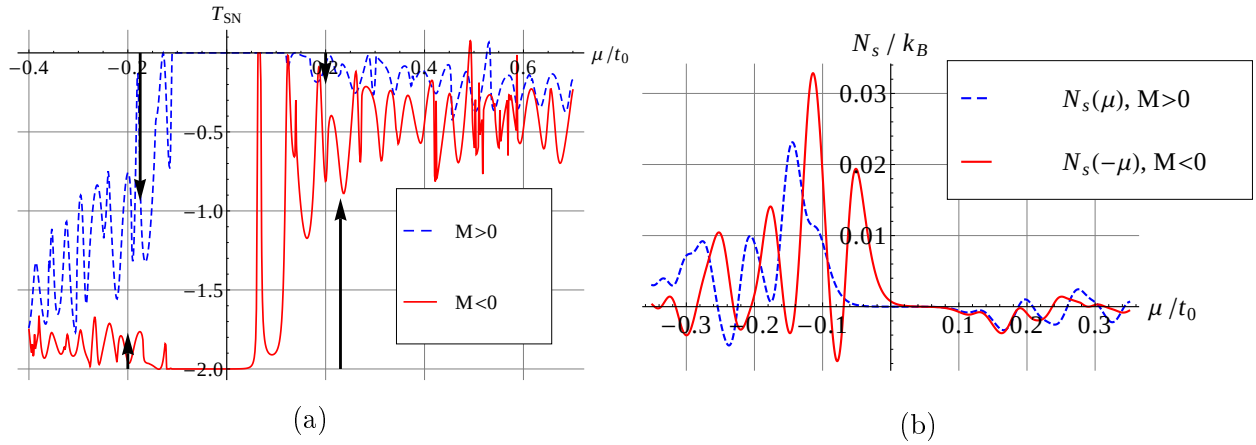


Figure 9.9: (a) Spin Nernst transmission function $\mathcal{T}_{SN}(\mu)$ for the normal regime ($M = 0.1t_0$, blue dashed) and inverted regime ($M = -0.1t_0$, red solid). In the bulk gap, the edge states give rise to a quantized spin Hall conductance of $e/2\pi$, except for the mini-gap. The black arrows are scaled according to the expected factor of $\frac{-m_h}{m_e} \approx 4.7$ between conduction and valence band spin Hall effect (see text). (b) Spin Nernst effect at T = 5 K, for normal (blue dashed) and inverted regime (red solid). The latter is flipped horizontally, so we can compare signals of the same band character. Taken from [Rot+12]. © (2012) by the American Physical Society

with $\Delta\phi(\theta) = 2\pi - \Delta\phi(-\theta)$. For not too small parameters A and small θ, the constant phase shift $\Delta\phi(0) = \pi$ is dominant. This phase shift ensures that the sign of the spin current is well-defined over a large range of θ. Therefore, even the superposition of many incident modes at different angles (not included in this simple analysis) would lead to a well-defined sign of the spin current near the interface, while far from the interface, the spin current will be suppressed by the oscillations.

9.4.3 Spin-Nernst signal for the bulk metallic regime

In the preceding sections, we showed that the spin current can be understood by an anomalous velocity or a spin-dependent phase shift. The expressions we have obtained do not depend on the effective band mass (considering the lowest order in Ak/M). We will now show that such a scaling of the spin current leads to a spin Nernst signal proportional to the effective band mass.

Let us assume that the applied difference in the chemical potential $\Delta\mu$ generates the spin (G_{sH}) and the charge (G_{xx}) responses in the system. Then $G_{sH} = \frac{I_s}{I} G_{xx}$. For a given number of modes G_{xx} is approximately constant. Therefore, using Eq. (9.31), one can see that $G_{sH} \sim |m_{e/h}|A^2/M$, where $m_{e/h}$ is the mass of the electron/ heavy hole band, respectively.[2] To the lowest order in Ak/M, the effective two-band and four-band masses coincide and the four-band effective masses $m_{e/h}$ are given by $\hbar^2/2m_e = -D - B$ and $\hbar^2/2m_h = -D + B$. Correspondingly, through the relation between the spin Hall conductance and the spin Nernst transmission signal $\mathcal{T}_{SN} \sim G_{sH} \sim |m_{e/h}|$. The last dependence can be easily seen in the limit for $D = 0$, when the band structure of the BHZ model is particle-hole symmetric. Then, $m_e = -m_h$ which is consistent with $\mathcal{T}_{SN} \sim |m_{e/h}|$ and the symmetry relation $\mathcal{T}_{SN}(\mu) = \mathcal{T}_{SN}(-\mu)$ in that case.

[2] For fixed electrical bias $\Delta\mu$, the current direction is fixed. Therefore, we have $I \propto 1/|m|$, even if flipping the sign of the effective mass m corresponds to changing electrons to holes.

Figure 9.9 (a) shows numerical results for $\mathcal{T}_{SN}$ as a function of the chemical potential. In the four-band model, the ratio of valence and conduction band effective masses is $\frac{-m_h}{m_e} \approx 4.7$. The black arrows are drawn for comparison of $\mathcal{T}_{SN}$ in conduction and valence band and are scaled by the factor $-m_h/m_e$. Their position is chosen for energies corresponding to 4 propagating modes in the leads (counting spin), not counting edge states. For the normal regime (see dashed lines in Fig. 9.9 (a)) the scaling of the numerical $\mathcal{T}_{SN}$ is very close to what we predicted from the analytical approaches. In Fig. 9.9(b), we show the corresponding spin Nernst signal. In the normal regime, we qualitatively find $N_s \propto |m_i|$ $(i = e, h)$ as expected from the Mott-like relation in combination with Fig. 9.9(a).

In the inverted regime (solid red line) we must consider that near the bulk gap, the band character (E/H) has changed (compare the red/blue colouring in Fig. 9.2); therefore, the band for $\mu > 0$ gets a heavy hole character. Further, as long as the edge states do not yet merge to the bulk, they are responsible for an offset of $\mathcal{T}_{SN} = -2$. The black arrows again indicate the factor $\frac{-m_h}{m_e}$ that we expect for the comparison of conduction and valence band signals at the same number of contributing modes, however now we are measuring the signal from the level of the edge states. Analysing numerically the scattering matrix we find that in the valence band the contribution to $\mathcal{T}_{SN}$ of bulk and edge states are additive, while this is not the case for the conduction band. Taking into account this fact, it is surprising that the simple analytical analysis applicable to the normal regime still describes qualitatively the numerics. We believe that this might be the case, because the first bulk state resembles the edge state character, and our argument about the symmetry of $\mathcal{T}_{SN}(\mu) = \mathcal{T}_{SN}(-\mu)$ for the particle-hole symmetric Hamiltonian still holds.

9.5 Conclusions

We have analysed the thermoelectric transport in four-terminal setups of HgTe/CdTe quantum wells with a particular emphasis on spin-dependent effects due to spin-orbit coupling. Thereby, we have used a combination of analytical and numerical methods to analyse spin-dependent transport phenomena. The Seebeck and the spin Nernst signal show a peculiar dependence on the parameters of the Bernevig-Hughes-Zhang model which can be qualitatively understood as originating from a spin Hall effect that arises at in-plane potential or confinement boundaries of the system. We have demonstrated that the spin Nernst effect is a strong experimental tool to get a better understanding of the mini-gaps that arise due to the spatial overlap of edge states on opposite sample boundaries. Most interestingly, we have derived a Mott-like relation between the spin Nernst coefficient and a smoothed spin Nernst transmission function that is valid to all orders in the temperature difference between the warm and the cold reservoir. Our findings might help to optimize future experiments on thermoelectric transport properties of two-dimensional topological insulators.

Chapter 10

Tunable polarization in beam-splitter based on 2D topological insulators

The bulk model describing 2D topological insulators (TI) derived in Chapter 7 contains two types of spin-orbit terms, the so-called Dirac terms which induce out-of plane spin polarization and the Rashba terms which can induce in-plane spin polarization. We show that for some parameters of the Fermi energy, a beam splitter device built on 2D TIs can achieve higher in-plane spin polarization than one built on materials described by the Rashba model itself. Further, due to high tunability of the electron density and the asymmetry of the quantum well, spin polarization in different directions can be obtained. While in the normal (topologically trivial) regime the in-plane spin polarization dominates, in the inverted regime the out-of plane polarization is more significant not only in the band gap but also for small Fermi energies above the gap. Further, we suggest a double beam splitter scheme, to measure in-plane spin current all electrically. Instead of directly analysing physical spin signals, we introduce polarization signals which are closer related to conserved quantities of a free system, in particular the helicity. We also discuss their relation to physical spin observables. This chapter is based on our publication [Rot+14].

10.1 Introduction and overview

One of the major goals in the field of spintronics is the generation and detection of spin currents [Wol+01]. Most of the proposals rely on external magnetic fields or ferromagnetic constituents. However, spin injection from a ferromagnet into a semiconductor turns out to be problematic. Attempts to remove the Schottky barrier by doping may lead to spin dephasing at the interface, and also conductivity mismatch turns out to be a fundamental problem [Sch+00] for spin current injection. There have been successful experimental attempts using vacuum tunnelling [Alv+92] and ferromagnetic semiconductors [Ohn+99]. Considering the impediments integrating ferromagnet-semiconductor interfaces or applied magnetic fields into the technology, it is desirable to find spin current injection based on all-electrical principles.

One possibility to generate a spin current by all electrical means is by using the spin-Hall effect (SHE), i.e. the generation of a spin current perpendicular to the applied field in the medium with a spin-orbit coupling. One of the sources of a transverse spin current in this context could be a spin-dependent scattering off impurities [Dya+71; Hir99]. Later, it has been realized that the spin Hall effect is also present in semiconductors with large spin-orbit (SO) induced band

splitting [Mur+03; Sin+04]. Spin accumulation induced by the spin-Hall mechanism has been experimentally confirmed first by optical means [Kat+04; Wun+05], and later, the SHE has been also all-electrically detected by combining it with the inverse SHE in metals and semiconductors [Val+06; Brü+10]. On the theoretical side, the interplay of different mechanisms leading to the SHE - scattering off impurities causing the skew scattering and side jump effects (so called extrinsic mechanisms), and SO splitting from the band structure (intrinsic effects) - has been a tough problem. The long standing debate has been only recently resolved showing that in the DC limit for the linear Rashba model the SHE due to the skew scattering and side-jump effects vanishes in the presence of the linear band-structure SO coupling while it is non-zero for the 2DEG with the dominant cubic Dresselhaus term [Tse+06; Han+08b; Han+09b; Bi+13].

All-electric generation of spin currents has also been proposed by pumping techniques, either by control over the Rashba coupling constant [Gov+03], or directly by gates controlling tunnelling constants [Bro+10a]. Spin pumping, however relying on Zeeman splitting, has also been experimentally proven [Wat+03].

Here, we follow a different idea, which has originally been proposed by Khodas [Kho+04]. Similar to birefringence in optics, different spin components are refracted differently at an interface where the Rashba SO coupling is non-zero on only one side - thus allowing for a construction of a spin filter. Based on this principle, there have also been proposals using graphene as material, which is interesting for the study of electron optics because of its linear dispersion. Bercioux et al. [Ber+10] have analysed spin-dependent transmission through an infinite N-SO-N (normal-Rashba SO - normal) junction based on graphene, and have also predicted spin pumping based on this geometry [Ber+12]. Further, SO terms in graphene have also been exploited to obtain spin-dependent Veselago lensing, allowing to focus the spin [Asm+13].

Differently from these works, we choose the HgTe/CdTe quantum well (QW) as a material of interest, because of its huge intrinsic Rashba SO coupling which can be electrically tuned. This material has also attracted a lot of interest because of its topologically non-trivial band structure. In this chapter, we compare polarization signals generated in the topologically trivial and topologically non-trivial regimes. We use numerical simulations to analyse generation of spin polarization and spin currents, and detection of the latter. However, instead of directly calculating spin polarizations and currents we rather choose to calculate transport of a conserved quantum number which is the helicity, and will later discuss the relation to the physical spin. We consider three different setups: the first one is an infinite interface of normal / Rashba SO systems, where the normal part is described by a Dirac-like Hamiltonian. The other two setups are very different from those of Khodas and Bercioux, since we consider realistic devices attached to leads. We consider a finite four-terminal beam splitter device, where generated spin polarization could be experimentally detected by Faraday or Kerr rotation, and also a double beam splitter device which is the most relevant experimentally, since it allows for all-electrical detection of the induced in-plane spin polarization (see Section 10.5).

Due to the Dirac-like form of the Hamiltonian for HgTe QWs [Ber+06a], even for zero Rashba terms, there is an intrinsic SO coupling already in the model leading to an out-of plane spin polarization (compare Chapter 7). Therefore, the physics is more complicated in our system, and shows a competition of different SO terms. Interestingly, for well chosen parameters, this can even increase the achieved in-plane spin polarization. While in the normal (topologically trivial) regime the in-plane spin polarization dominates, in the inverted regime the out-of plane polarization is more significant not only in the band gap but also for small Fermi energies above the gap.

The outline of this chapter is as follows. In the Section 10.2, we present the model Hamiltonian for HgTe QW and show how we define and characterize the helicity polarization that is induced by Rashba SO. In Section 10.3, we discuss the simple N-SO interface, where N is defined by the Dirac-like model [Ber+06a] and SO denotes the Dirac-like model with linear and cubic Rashba SO interactions. Since this problem corresponds to a matrix-valued third order differential equation for the envelope function, there are issues with the proper definition of currents [Li+07], proper boundary conditions [Win+93], and spurious solutions [Sch+85]. We find that mapping the problem to a lattice, gives an elegant solution to all these problems. This approach is similar to an idea by Winkler [Win+93]. We give some technical details about our wave matching approach in Appendix F.

In Section 10.4, we consider an N-SO-N junction embedded in a realistic four-terminal device. Here we discuss only a linear Rashba term in the SO region, as is justified on the basis of a single junction analysis. We discuss aspects about a geometry suitable for splitting of helicity components, discuss numerical methods, and also elaborate on some technical questions concerning the measurement of spin currents or helicity currents. Finally, to obtain a better understanding of the interplay between an in-plane polarization coming from Rashba physics and an out-of plane polarization originating from the Dirac-like physics, we employ an effective perturbative model that has also been used in Chapters 7 and 9.

To link the numerical analysis with possibilities of experimental detection, we first discuss a combined device with polarizer and analyzer, in Section 10.5, and obtain a relation between helicity and the physical electron spin, in Section 10.6.

In Appendix C, we give a detailed derivation and discussion of the helicity operator for the 4-band model. Details on the S-matrix and how it is related to the helicity current are given in Appendix D. Measurement and correct definition of spin currents in structures, where spin is not conserved, has raised a lot of discussion in the literature. Therefore, we derive a generalized continuity equation including a torque term [Shi+06], adapted to our Hamiltonian, in Appendix E. This serves to underline the physical meaning of the helicity current.

10.2 Model and characterization of polarization

We consider the 4-band model for the lowest subbands of a CdTe/HgTe/CdTe quantum well (QW), written in the basis $|E+\rangle$, $|H+\rangle$, $|E-\rangle$, $|H-\rangle$ of electron-like and heavy-hole QW subbands, which are angular momentum eigenstates with $S_z = \frac{1}{2}, \frac{3}{2}, -\frac{1}{2}, -\frac{3}{2}$. The Hamiltonian [Rot+10] is an extension of the Bernevig-Hughes-Zhang model [Ber+06a; Kön+08], and details on the derivation can be found in Chapter 7.1.1.

$$H(\mathbf{k}) = \epsilon(k)I + \begin{pmatrix} \mathcal{M}(k) & Ak_+ & -iRk_- & -iSk_-^2 \\ Ak_- & -\mathcal{M}(k) & iSk_-^2 & iTk_-^3 \\ iRk_+ & -iSk_+^2 & \mathcal{M}(k) & -Ak_- \\ iSk_+^2 & -iTk_+^3 & -Ak_+ & -\mathcal{M}(k) \end{pmatrix} \tag{10.1}$$

with $\mathcal{M}(k) = M - Bk^2$, $\epsilon(k) = C - Dk^2$, $k^2 = k_x^2 + k_y^2$ and $k_\pm = k_x \pm ik_y$. The parameter M describes the band gap and is tunable by the QW width d with $M > 0$ for the trivial insulator and $M < 0$ for the topological insulator. We take the parameters $A = 0.365$ eV nm, $B = -0.50$ eV nm^2, $D = -0.50$ eV nm^2 and $M = 24$ meV, corresponding to a realistic experimental situation [Müh+14b] with $d = 5$ nm, in the trivial insulating regime. For the inverted regime, we take band parameters $A = 0.375$ eV nm, $B = -1.120$ eV nm^2, $D = -0.730$ eV nm^2 and

$M = -10$ meV, corresponding to a QW width $d = 7$ nm. We will work with the parameters for the normal regime most of the time, comparing with the inverted regime in the Section 10.4.

We can decompose

$$H(\mathbf{k}) = H_N(\mathbf{k}) + H_{SO}(\mathbf{k}), \tag{10.2}$$

where H_N contains the two diagonal 2×2 blocks and H_{SO} consists of the two off-diagonal 2×2 blocks, which depend on the Rashba SO parameters R, S and T. The latter are tunable by the asymmetry of the QW, e.g. by top or bottom gates. The parameters S and T will give only small corrections to conduction band properties compared to R. For a given perpendicular electric field, we take the ratios S/R and T/R from our earlier work [Rot+10]. The Fermi energy E_F is also experimentally tunable by top or bottom gates. We will only consider E_F lying in the electron-like band. In the following $H(\mathbf{k})$ means the 4×4 Hamiltonian matrix, and $\hat{H} = H(\hat{\mathbf{k}})$ means its real space or lattice representation.

We introduce a generalization of the helicity operator to the 4-band model,

$$h(\mathbf{k}) = \begin{pmatrix} 0 & 0 & -ik_-/k & 0 \\ 0 & 0 & 0 & i(k_-/k)^3 \\ ik_+/k & 0 & 0 & 0 \\ 0 & -i(k_+/k)^3 & 0 & 0 \end{pmatrix}. \tag{10.3}$$

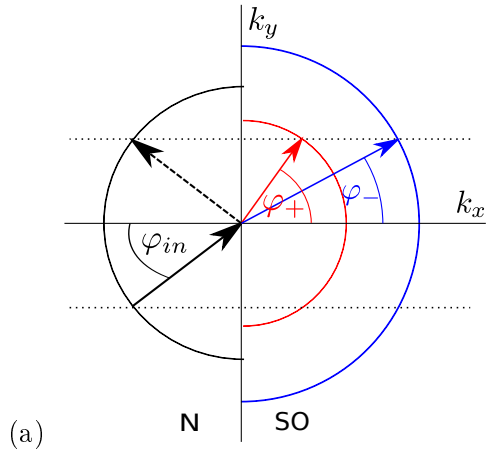
$\hat{h} = h(\hat{\mathbf{k}})$ has the same symmetries as the conventional helicity operator $\boldsymbol{\sigma}\mathbf{p}/p$ of the Pauli equation, i.e. it is parity-odd, time-reversal even, and $[\hat{H}, \hat{h}] = 0$. The eigenvalues ± 1 of $h(\mathbf{k})$ are two-fold degenerate. We will use local expectation values of $h(\mathbf{k})$ to define a "helicity polarization" which is analogue to an in-plane spin polarization. At a given direction $\mathbf{k}/k$ and given E_F, there are two propagating eigenstates. Let χ be the normalized 4-component spinor part. The propagating state can be chosen such that χ diagonalizes $h(\mathbf{k})$, i.e. $\chi^\dagger h(\mathbf{k})\chi$ will be ± 1. However, $\mathbf{k}/k$ will not be always well-defined, but rather approximately known, by the direction of a lead that guides an electron beam. Therefore, we just fix the direction $\mathbf{k}/k$ and define $h_x = h(k = k_x) = \tau_z\sigma_y$, where σ Pauli matrices act on $(+,-)$-space, τ Pauli matrices act on (E,H)-space, and σ_0, τ_0 denote unit matrices. We then use the local expectation value of h_x for characterization of the helicity. For a detailed discussion of the helicity operator, see Appendix C.

The local spin-z polarization, on the other hand, will be measured by the operator $\tau_0\sigma_z$ on the band space (note, it is not S_z). If Rashba SO terms are zero ($R = S = T = 0$), then $[\hat{H}, \tau_0\sigma_z] = 0$.

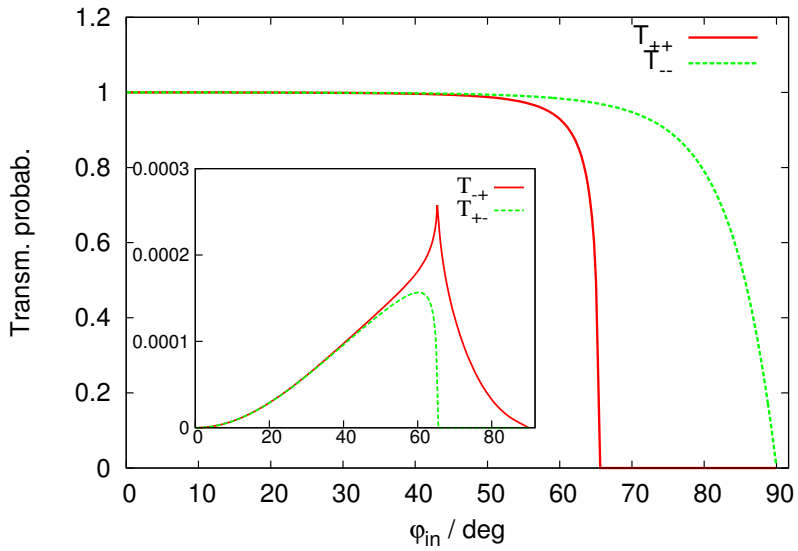
10.3 Infinite N-SO interface

In this section, we search for a maximal value of the helicity polarization at an infinite N-SO (normal - Rashba spin orbit) interface, by means of total reflection. We show that transmission from a mode of one helicity into a mode of the opposite helicity (a cross-helicity term) is very small.

We consider the simple N-SO setup of Fig. 10.1a. To the left of the interface at $x = 0$, we have vanishing Rashba SO coupling terms (Hamiltonian given only by $\hat{H}_N$ of Eq. 10.2), and to the right we assume constant non-zero values for the Rashba parameters R, S, T (full $H(\mathbf{k})$ in Eq. 10.2). An electron beam enters from the left, with angle of incidence φ_{in}. To the right of the interface, the components of different helicity (± 1) continue at different angles $\varphi_\pm$.



(a)



(b)

Figure 10.1: (a) Refraction of an electron beam at an infinite N-SO interface, where the Rashba spin-orbit Hamiltonian is given by $H_{SO}(\mathbf{k})$. For the same incoming angle φ_{in}, the outgoing angle $\varphi_\pm$ depends on the helicity ± 1. The black semicircle on the left indicates the Fermi energy condition $E_0(k) = E_F$ in momentum space, for zero Rashba SO coupling, while the red/blue semicircles on the right indicate the conditions $E_\pm(k) = E_F$. The dotted horizontal lines correspond to the conserved momentum k_y. (b) The transmission probabilities $T_{\sigma,\sigma'}$ between different helicities σ, σ' as a function of the angle of incidence onto the infinite interface. Here the Fermi energy is $E_F = 13.8$ meV, the Rashba SO parameters are $R = 40$ meV nm, $S = 5.4$ meV nm^2 and $T = 23$ meV nm^3 and a lattice constant of $a = 4.94$ nm is used. Figure (b) is taken from [Rot+14]. © (2014) by the American Physical Society

The N-SO interfaces will be the building block of the more realistic setup of the next section, compare Fig. 10.2. Compared to the analysis of Khodas [Kho+04] who also considered an N-SO interface, our model Hamiltonian is more complicated because it contains more bands and in particular, non-linear Rashba SO terms. Our approach can also be used for SO-SO interfaces with different Rashba SO parameters to the left and right, but for clarity, we restrict ourselves to the N-SO case.

The model Hamiltonian is

$$\hat{H} = \hat{H}_N + \theta(x)\hat{H}_{SO}\theta(x) \tag{10.4}$$

with the Heaviside function $\theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$. The exact form of the symmetrization of the non-commuting parts $\theta(x)$ and momentum-dependent $\hat{H}_{SO}$ will have influence on the sharpness of the interface on the scale of a lattice constant. However, this is a detail that has no strong influence on the resulting transmission coefficients. The chosen symmetrization gives the sharpest possible interface, sharper than e.g. $\frac{1}{2}\left\{\theta(x), \hat{H}_{SO}\right\}$.

The outgoing beam directions and the critical angle of total reflection of the + helicity component are all fixed by the conserved quantities E_F and k_y. It is easy to show that at the interface, cross-helicity transmission and reflection probabilities are zero for waves entering perpendicular to the interface ($\varphi_{in} = 0$). (For this, we show that in the subspace of eigenstates with $k_y = 0$, the space dependent Hamiltonian $\hat{H}$ and $\hat{h}$ can be diagonalized simultaneously). For a given Fermi energy, the critical angle φ_c can be already found from the dispersions, without performing any wave matching. Let's call $E_+(k)$ the dispersion of the SO region with positive helicity, and $E_0(k)$ the dispersion of the N region. Finding k_y from $E_+(k_y) = E_F$ and solving for $k_{0,x}$ in $E_0(\sqrt{k_{0,x}^2 + k_y^2}) = E_F$, we have $\tan\varphi_c = \frac{k_y}{k_{0,x}}$. We note in passing, that a top gate which gives a potential step at $x = 0$, could be used to modify the critical angle, but we will not make use of this option here. In particular we want to see if cross-helicity transmission probabilities $T_{\pm\mp}$ are also small for incoming angles $\varphi_{in} \neq 0$. The difficulty is, that upon replacing $k_x \to -i\partial_x$, we obtain a differential equation of third order, and exact matching conditions at the boundary $x = 0$ are difficult to find. Instead, we use a lattice approximation of $\hat{H}$, with next nearest neighbour couplings, so we can match the dispersion of the analytical model to Fourier components $\sin(ak_x)$, $\sin(2ak_x)$, $\cos(ak_x)$ and $\cos(2ak_x)$, which enter the dispersion of the lattice model. Moreover, this method also automatically excludes large-k spurious solutions [Sch+85; Win+93] that appear for the third order differential equation, compare also the discussion in Chapter 2.4. This approximation breaks rotational symmetry, but we still find quite good helicity values (the expectation value is different from ± 1 by less than 10^{-6}). Further, this method is also quite flexible, e.g. one may smoothen the interface in the model. Transmission and reflection coefficients are then calculated with the equilibrium Green's function method. For a detailed discussion of the wave matching method, we refer to Appendix F.

Figure 10.1b shows the transmission probability as a function of φ_{in}. We see that total reflection can occur for the + component. (We assume $R, S, T \geq 0$, so the + component is the one with the higher energy for the same k. If we switch the sign of the perpendicular electrical field, R, S, T change signs and total reflection will occur for the − component.) For $\varphi_{in} \neq 0$, helicity is not conserved, but the non-conservation, given by the cross-helicity transmissions $T_{\pm\mp}$, is quite small. For comparison, we have carried out the same calculation, keeping only

linear Rashba terms, and we find that the critical angle is only slightly changed, while the cross - helicity transmission is an order of magnitude smaller, and thus also negligible. We conclude that we do not need to include other than linear Rashba terms in the following section, where we present calculations for a realistic geometry.

Khodas [Kho+04] also discusses the case of a smooth interface, with the Rashba SO coupling being turned on smoothly over a distance d_R. If d_R is large compared to the Fermi wavelength of the incident electron beam, the spinor adiabatically adjusts and the signals T_{++} and T_{--} in Fig. 10.1b become step functions, while $T_{\pm\mp}$ will be suppressed completely. However, since our (and Khodas') proposal relies on helicity (or spin) filtering by means of total reflection and conservation of k_y is not affected by the smoothness of the interface, the scale d_R is not important for the operation of the helicity (or spin) polariser.

10.4 Polarization in finite systems

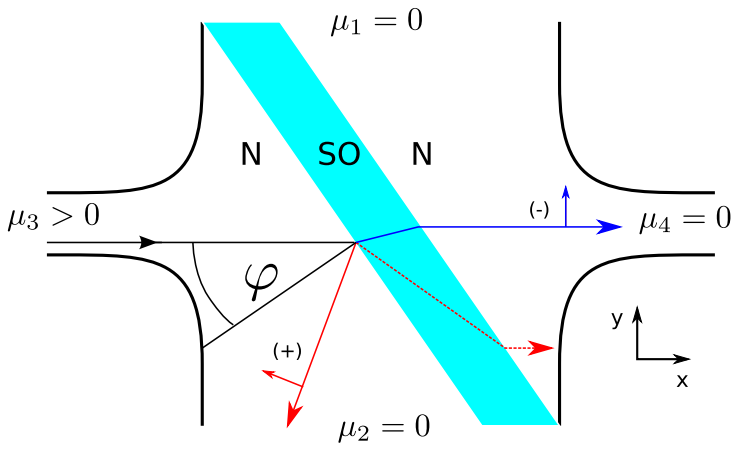
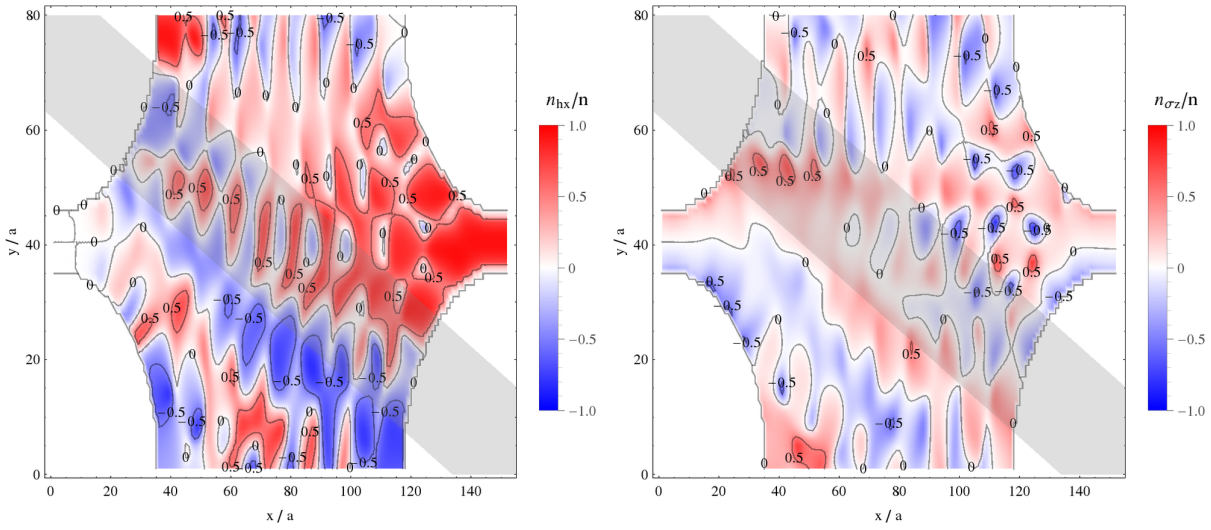


Figure 10.2: The beam splitter consists of a N-SO-N junction, tilted at angle φ, which is embedded in a four-lead device. If $\varphi > \varphi_c$, we expect total reflection of one helicity component at the first N-SO interface, causing the beam leaving through the lead 4 to be polarized. The dotted red line shows the case when $\varphi < \varphi_c$. Taken from [Rot+14]. © (2014) by the American Physical Society

In this section, we only consider the linear Rashba coupling R, because it is the most relevant, and put $S = T = 0$. In contrast to the N-SO interface discussed before and also analysed by Khodas, we are looking for a good implementation of a spin or helicity filter in a finite geometry with attached leads. We give some thoughts on the geometry in the Subsection 10.4.1, and recognize that a good spin filter device will have the form of the four-lead setup shown in Fig. 10.2. Next, in the Subsection 10.4.2, we give details about the numerical methods used to obtain polarization and current signals, and show results. We give a numerical comparison of the average helicity density and average helicity current in the leads and show (in Appendix E) that they are linked by a continuity equation. Finally in the Subsection 10.4.3, we provide a better understanding of the numerical results, employing an effective 2-band model. In particular, this helps to understand the competition of the different SO terms present in the 4-band model, which cause in-plane and out-of plane spin polarization. We shortly comment on the validity of



(a) (b)

Figure 10.3: Local polarization plots, obtained by discretizing the 4-band model Hamiltonian on a lattice (700 nm x 400 nm, lattice constant $a = 4.94$ nm), and a geometry corresponding to Fig. 10.2. Band parameters are for the normal regime. We show the non-equilibrium response to a bias applied at the left lead. The Fermi energy is $E_F = 0.337t_0 = 13.8$ meV, corresponding to a peak in the polarization of the outgoing helicity current (right lead). The linear SO parameter R is non-zero only in the opaque gray area, which has a horizontal extension of $54a$ and which is tilted at an angle of $\varphi = 65°$, which is approximately equal to the critical angle φ_c at the chosen E_F. The value of R is given by $t_{SO}/t_0 = 0.1$. (a) shows the local normalized helicity polarization assuming the direction $k = k_x$, $n_{hx}(\mathbf{r})/n(\mathbf{r})$. (b) shows the normalized σ_z-polarization $n_{\sigma z}(\mathbf{r})/n(\mathbf{r})$, which originates mainly from the Dirac-like physics. Taken from [Rot+14]. © (2014) by the American Physical Society

the model in Subsection 10.4.4.

10.4.1 Setup geometry

In a realistic electronic micro-device, the N-SO interfaces will have finite extension and the in/outgoing electron beam will be guided by the attached leads. Figure 10.2 shows a tilted N-SO-N junction embedded in a four-lead device. The electrical boundary conditions, i.e. the applied potentials μ_i at the leads, are such that an electron beam enters from the left, at an angle of incidence φ to the N-SO surface. φ should be above the critical angle of total reflection φ_c of the + component. Then only the − component traverses the SO barrier and leaves it in the same direction as it has entered. If $\varphi < \varphi_c$, we can still expect some helicity polarization in the right lead, because of the parallel offset of the passing + beam.

The left and right leads are narrow (49 nm width) and widen slowly (adiabatically). This ensures that by the horn collimation effect (see e.g. [Bee+91] for a quantum mechanical discussion of collimation), the beam injected from the lead is well-directed. The Fermi energy is chosen low enough to have only two propagating modes in the left and right leads (not counting edge states if we are working in the inverted regime). In an experimental setup, this collimation can be achieved by quantum point contacts. The upper/lower leads have to be wide to reduce

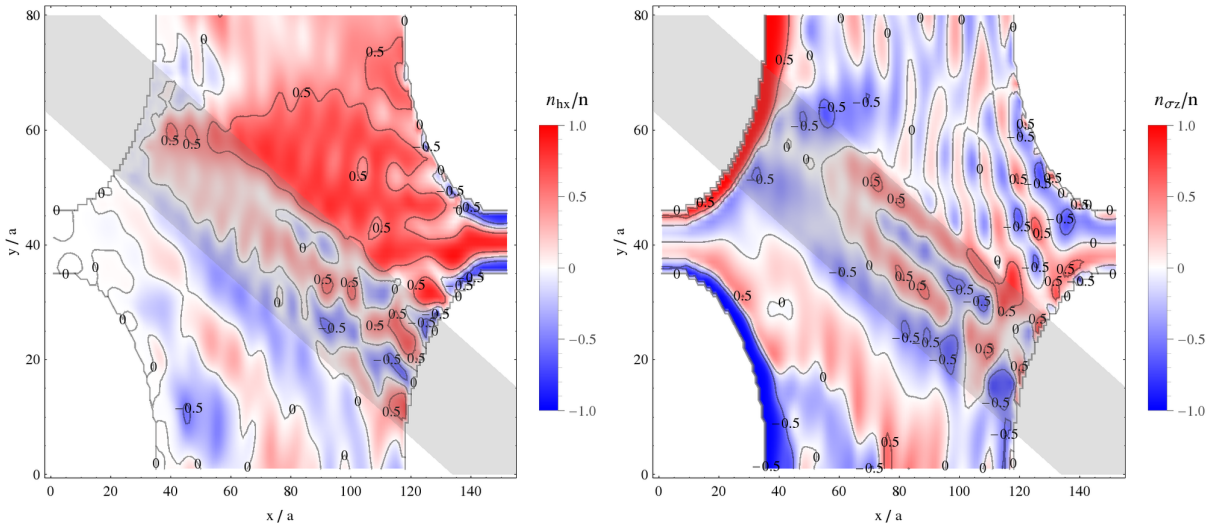


Figure 10.4: Similarly to Figs. 10.3, the local normalized helicity polarization (a) and σ_z-polarization (b) are shown, but for the inverted ($M < 0$) regime. The bulk gap lies in the range of $[-10, 10]$ meV. We choose $E_F = -17.6$ meV inside the (electron-like) valence band, corresponding to 4 propagating modes in the left/right leads. The linear Rashba coupling in the barrier is again fixed by choosing $t_{SO}/t_0 = 0.1$. Taken from [Rot+14]. © (2014) by the American Physical Society

undesirable reflections.

Our setup is invariant under rotation by π about $\hat{z}$ (C_2 symmetry). Thus, a beam passing the device from the left will be polarized with the same efficiency as when it enters from the right. A simple argument shows that this symmetry is required to achieve efficient filtering of the helicity. A device that has efficient spin/helicity filtering but not conversion, will have $T_{3\sigma,4\bar{\sigma}} \approx 0$, where the $\bar{\sigma}$ denotes $-\sigma$, and $T_{p\sigma,q\sigma'}$ denotes the transmission probability from lead q considering only modes with helicity σ', to lead p, with only modes with helicity σ. Here, we use the word "helicity" in a loose way, assigning helicity + to a transverse mode χ if the expectation value $\chi^\dagger h_x \chi$ is positive. It will be less than 1 because the transverse modes are not $\hat{k}_y$-eigenstates with $k_y = 0$, but still, the sign is sufficient to find a simple description of the time reversal symmetry of the S-matrix. If we sort in- and outgoing modes according to their helicities, time reversal symmetry makes the matrix of transmission probabilities symmetric, $T^T = T$. Then, $\Delta T_{4,3} := T_{4+,3+} + T_{4+,3-} - T_{4-,3+} - T_{4-,3-} \approx T_{4+,3+} - T_{4-,3-} \approx \Delta T_{3,4}$. A good efficiency in polarizing a beam entering at lead 3 means $|\Delta T_{4,3}|$ should be large, and we see that this requires that a beam entering at lead 4 and exiting at lead 3 has to be also well polarized. With C_2 symmetry, we indeed have $T_{4+,3-} \overset{C_2}{=} T_{3+,4-} \overset{TR}{=} T_{4-,3+}$ and thus $\Delta T_{4,3} = \Delta T_{3,4}$ holds exactly.

Khodas [Kho+04] also proposes spin filtering at a single interface, based on the outgoing angle. Compared to the N-SO-N interface employed here, spin filtering then works in a wider range of incoming angles, since the angle of incidence may be less than the critical angle. However, if leads are attached to collect the + and − components at different angles, this will break the C_2 symmetry. Further, the outgoing beam would need re-collimation. In contrast, in our setup the outgoing beam has the same direction as the incoming beam, and attaching leads keeps the C_2

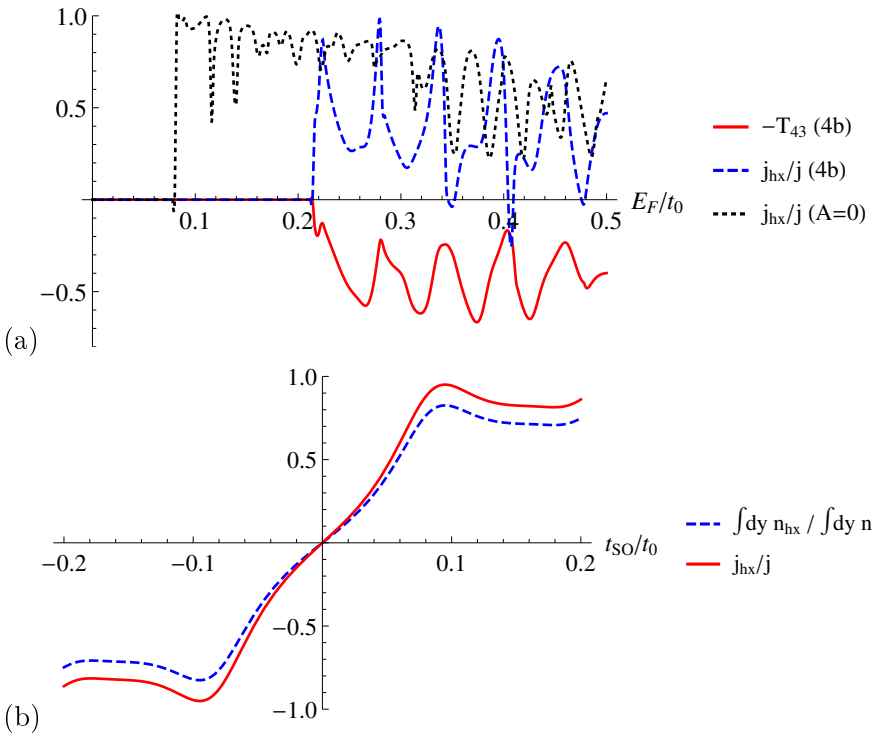


Figure 10.5: (a) The normalized helicity current j_{hx}/j at the right lead as function of the Fermi energy is shown, for the normal regime. Its third peak lies at $E_F = 0.337t_0$ ("sweet spot"). For comparison, j_{hx}/j is also plotted for a 2-band model obtained by setting $A = 0$, where it reduces to the normalized spin-y current. The outgoing current at the right lead, which is proportional to T_{43}, is shown with flipped sign for clarity. (b) Comparison of normalized helicity current and normalized average helicity polarization $\int dy\, n_{hx} / \int dy\, n$ in the right lead. Taken from [Rot+14]. © (2014) by the American Physical Society

symmetry.

Note also, that with time reversal invariance and current conservation, a two-lead spin filter is impossible with only two propagating modes (including spin degeneracy), so the extra leads are required. This is because, analogous to Kramers degeneracy, the eigenvalues of the matrix $tt^\dagger$, which gives the transmission probabilities, are two-fold degenerate [Bar08]. Here, $t = S_{a,b}$ is a 2×2 submatrix of the 4×4 S-matrix, for 2 modes in each of the leads a,b. Thus $tt^\dagger = \gamma \mathbf{1}$ is proportional to the unit matrix ($\gamma \in \mathbb{R}$). Introducing the projector P_σ on some unspecified spin direction, we find $T_{b\sigma,a} = \mathrm{Tr}[t^\dagger P_\sigma t] = \gamma \mathrm{Tr}[P_\sigma] = \gamma$. The result is independent of the spin direction, thus making filtering of the spin impossible.

10.4.2 Formalism and polarization/current signals

We are interested in the helicity current at the right lead, due to the applied bias μ_3 at the left lead. To locally investigate some operator $\hat{O}(\mathbf{r})$ describing the polarization or current at position $\mathbf{r} = (x, y)$, we can plot

$$\frac{\partial \langle \hat{O} \rangle(\mathbf{r})}{\partial \mu_3} = \frac{1}{2\pi} \mathrm{Tr}\left[A_3(E_F) \hat{O}(\mathbf{r}) \right] \tag{10.5}$$

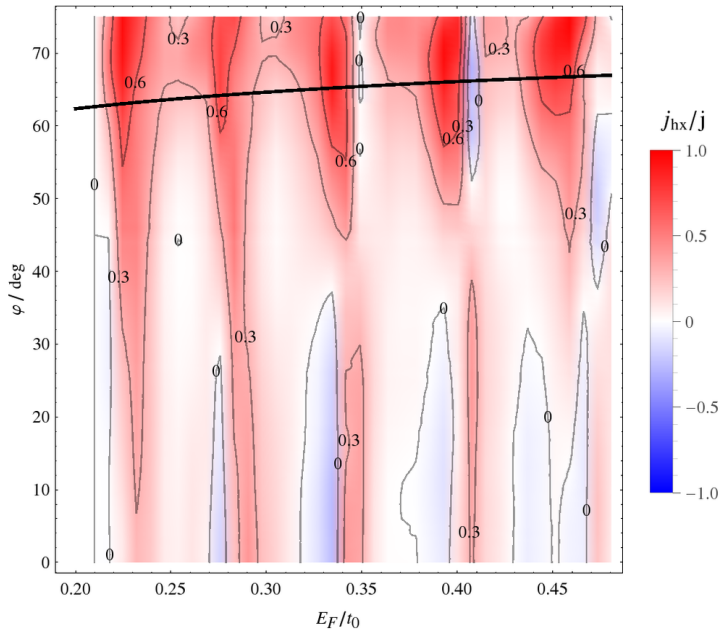


Figure 10.6: Normalized helicity current as function of Fermi energy and beam splitter tilting angle φ, for the normal regime. Non-zero helicity current is already visible when $\varphi = 0$. However, good polarization is only seen for $\varphi \approx \varphi_c$ or greater. The thick line shows $\varphi_c(E_F)$. Taken from [Rot+14]. © (2014) by the American Physical Society

with the left lead contribution to the spectral density

$$A_3(E_F) = G^R \Gamma_3 G^A. \tag{10.6}$$

Here $G^R = (E_F - \hat{H} - \sum_p \Sigma_p)^{-1}$ is the retarded Green's function, $G^A = (G^R)^\dagger$ the advanced Green's function, and $\Gamma_3 = i(\Sigma_3 - \Sigma_3^\dagger)$ is obtained from the left lead self-energy. The retarded self-energy of lead p is given by $\Sigma_p = \tilde{\tau}_p (E_F + i0^+ - \hat{H}_p)^{-1} \tilde{\tau}_p^\dagger$, $\tilde{\tau}_p$ is the matrix connecting the surfaces of lead p and sample, and $\hat{H}_p$ is the Hamiltonian of the isolated semi-infinite lead.

For the operator $\hat{O}(\mathbf{r})$ we insert $P_\mathbf{r} h_x$ or $P_\mathbf{r} \sigma_z$ for analysis of the local helicity or spin-z polarization, respectively, where $P_\mathbf{r} = |\mathbf{r}\rangle\langle\mathbf{r}|$ is the projector on coordinate $\mathbf{r}$. Figure 10.3 shows 2D density plots of these signals (normalized by the local density).

For this purpose, the 4-band model Hamiltonian (10.1) is discretized on a 700 nm × 400 nm lattice with lattice constant $a = 4.94$ nm, and a geometry corresponding to Figure 10.2. The linear SO parameter R is non-zero only in the opaque gray area, which has a horizontal extension of $54a$ and which is tilted at an angle of $\varphi = 65°$. The value of R is given by $t_{SO}/t_0 = 0.1$, where $t_{SO} = \frac{R}{2a}$ is the energy scale of Rashba SO in the lattice model and $t_0 = \frac{-B-D}{a^2}$ is the hopping energy for quadratic terms.

For the normal regime, we choose a Fermi energy $E_F = 0.337 t_0 = 13.8$ meV, corresponding to a peak in the polarization of the outgoing helicity current (right lead). The critical angle of total reflection is energy dependent, see also Figure 10.6 where the thick black line shows $\varphi_c(E_F)$. For our choice of E_F, we have $\varphi \approx \varphi_c$.

Figs. 10.3a and 10.4a show the normalized local helicity polarization, for the topologically trivial and topologically non-trivial regimes, respectively. We assume the direction $k = k_x$, i.e.

we show the expectation value $n_{hx}(\mathbf{r}) = \frac{\partial\langle P_\mathbf{r} h_x\rangle}{\partial\mu_3} d\mu_3$, normalized by the local (non-equilibrium) density $n(\mathbf{r}) = \frac{\partial\langle P_\mathbf{r}\sigma_0\tau_0\rangle}{\partial\mu_3} d\mu_3$. We see that the beam is partly polarized after passing the SO barrier. Figs. 10.3b and 10.4b show the normalized σ_z-polarization, for the topologically trivial and topologically non-trivial regimes, respectively. We show $n_{\sigma z}(\mathbf{r}) = \frac{\partial\langle P_\mathbf{r}\sigma_z\tau_0\rangle}{\partial\mu_3} d\mu_3$, again normalized by the local density.

The local spin-z and spin-y polarizations could be measured experimentally by Faraday or Kerr rotation. However, to compare with an experiment, the physical spin polarization is relevant rather than the local expectation values of the operators σ_y and h_x. The relation to the physical spin is discussed in Section 10.6, and here we just give some numerical results, referring to the physical in-plane and out-of plane spin polarization as s_y-polarization and s_z-polarization. We consider the right lead (coordinate $x = 150a$), and find maximal values of the spin polarization along the spatial y-coordinate, for different spin polarization directions. In our scale, 0 stands for equal contributions of spin up and down, while ± 1 means full polarization. For the normal regime (Figs. 10.3a,b), we find 28% σ_z-polarization, corresponding to 27% s_z-polarization at the upper edge of the lead, and 94% h_x-polarization corresponding to 76% s_y-polarization for the center of the lead. For the inverted regime (Figs. 10.4a,b), we find up to 38% σ_z-polarization, corresponding to 36% s_z-polarization, and up to 83% h_x-polarization, corresponding to 64% s_y-polarization. Note that this is for the electron-like regime of the inverted regime, i.e. energies below the gap. For energies above the gap, physical in-plane spin polarization is strongly suppressed, since heavy-hole components of the wave function do not contribute to it (see also Eq. (10.12) and the text below it).

We could also insert for $\hat{O}(\mathbf{r})$ the helicity current or spin-z current operators, $\hat{J}_{hx}(\mathbf{r}) = \frac{1}{4i}\{P_\mathbf{r}, \{h_x, [\hat{x}, \hat{H}]\}\}$ and $\hat{J}_{\sigma z}(\mathbf{r}) = \frac{1}{4i}\{P_\mathbf{r}, \{\tau_0\sigma_z, [\hat{x}, \hat{H}]\}\}$. To underline the physical meaning of such currents in transport, we have derived a general continuity equation in Appendix E. It includes a torque term [Shi+06] acting as source. However, when the average over the semi-infinite lead is taken, the torque vanishes, while the helicity current remains. If we are interested in the signal only in the right lead, instead of calculating the full Green's function, it is more efficient to obtain the currents from the scattering matrix and expectation values of the relevant operator evaluated within the mode basis of the lead. This way, we obtain the spin or helicity current, averaged over the semi-infinite lead. The method also works if Rashba SO terms are non-zero in the leads. See Appendix D for details on the operator expectation values in terms of the S-matrix. The scattering matrix entries $t_{pn,qm}$, i.e. the transmission amplitudes from lead q, mode m to lead p, mode n, are calculated using a generalized Fisher-Lee relation [Wim08], see Eq. (F.3) of Appendix F. The particle current at the right lead is proportional to the transmission probability straight through device, $T_{43} = \sum_{n,m} |t_{4n,3m}|^2$. Actually, in this work we choose the Fermi energy low enough to have only two propagating modes (counting spin degeneracy), and since we do not include SO terms in the leads the helicity current in the latter is conserved and identical to its average. Note that we cannot apply the method that is usually used to calculate spin currents by introducing separate leads for both spin (or here, helicity) directions, because $[\hat{H}, h_x] \neq 0$.

Since the non-equilibrium helicity polarization and helicity current should be both generated by filtering out a mode of particular helicity, we expect signals to be qualitatively the same. We can confirm this by the plot shown in Figure 10.5b, which shows a comparison of normalized helicity current and helicity polarization. The normalized helicity current at the right lead is given by j_{hx}/j, with the definitions $j_{hx} = \int dy\, \frac{\partial\langle \hat{J}_{hx}(x_4,y)\rangle}{\partial\mu_3} d\mu_3$ and $j = \int dy\, \frac{\partial\langle \hat{J}(x_4,y)\rangle}{\partial\mu_3} d\mu_3 \propto T_{43} d\mu_3$, where x_4 is far in lead 4 and $\hat{J}(\mathbf{r})$ is defined like $\hat{J}_{hx}(\mathbf{r})$ with h_x replaced by the unit

matrix. The normalized average helicity polarization is given by $\int dy\, n_{hx}(x_4, y)/\int dy\, n(x_4, y)$. The signals do not oscillate as function of x_4, since we have only two propagating modes in the right lead, with the same k_x. We have checked that the values obtained from the full Green's function, Eq. (10.5) and values obtained from the S-matrix (Appendix D) which needs only the surface Green's function, are almost identical. This is because evanescent modes are unimportant in the leads. Figure 10.5a shows the normalized helicity current j_{hx}/j at the right lead as function of the Fermi energy (blue dashed line, for the 4-band model). The third peak lies at $E_F = 0.337t_0$. We will use this point in the following sections, referring to it as sweet spot. Also, Figures 10.3a,b are calculated for this energy. In Fig. 10.5a, we also plot the helicity-independent transmission probability T_{43}, with a flipped sign for clarity. We see that its absolute value decreases whenever the current becomes polarized (due to the fact that the electron beam is split, less electrons are transmitted to the lead 4).

In Fig. 10.6, we show the normalized helicity current as function of both Fermi energy and beam splitter tilting angle φ. Non-zero helicity current is already visible when $\varphi = 0$. This is allowed by symmetry, since we have more than two leads, but it is not an effect of helicity-dependent refraction. The critical angle $\varphi_c(E_F)$ is shown as thick black line. Good polarization is only obtained for $\varphi \approx \varphi_c$ or larger, where we can explain the signal by the parallel displacement or total reflection of the + beam component.

10.4.3 Effective 2-band model

In Fig. 10.3a the spatial map of helicity polarization $n_{hx}(\mathbf{r})/n(\mathbf{r})$ is presented (normal regime). One can see that helicity polarization oscillates as a function of the spatial coordinate. In Fig. 10.3b the σ_z-polarization is shown. One can see that there is a large polarization close to the sample boundaries, which is well visible near the left and right leads. Here it is important to mention that there is non-zero out-of plane polarization, even without the SO barrier, as a consequence of the Dirac physics. However, as will be discussed below, the SO barrier allows to tune the degree of this polarization. To get a better insight into these results, we will use an effective 2-band model for just the $|E\pm\rangle$ bands in the low energy limit $Ak \ll M$, that we already used in Chapters 7 and 9. Since this model is strictly valid in the normal regime, we will start with the discussion of this regime before we analyse the inverted (topologically non-trivial) regime.

Since electron components of the wave function are dominant, the helicity (h_x) polarization is approximately given by the σ_y-polarization which can be analysed in the effective 2-band model. To allow analytical treatment, the confinement of the electrons is modelled by a confinement potential $\tau_z V(\mathbf{r})$, which replaces the lattice truncation for the desired geometry. This corresponds to a space-dependent band gap, so that both conduction and valence band states are confined. The effective Hamiltonian obtained in 3rd order perturbation theory is

$$\hat{H}_e = \frac{k^2}{2m^*} + \underbrace{R(\boldsymbol{\sigma} \times \hat{\mathbf{k}})_z}_{\text{Rashba}} + \underbrace{\frac{A^2}{4M^2}(\nabla V \times \hat{\mathbf{k}})_z \sigma_z}_{\hat{H}_D} + V(\mathbf{r}) \tag{10.7}$$

with the renormalized effective mass $m^* = (-2B - 2D - \frac{A^2}{M})^{-1}$.

We expect to see a competition between the Rashba and Dirac physics in the beam splitter device. However, only the magnitude of the Rashba coupling is tunable by top and/or bottom gates. The Dirac physics, given by $\hat{H}_D$, is due to the intrinsic SO coupling of the HgTe/CdTe

material, and let us emphasize that this SO term is absent in the 2DEG model analysed by Khodas. In Figure 10.3b, the non-zero value of $n_{\sigma z}(\mathbf{r})$ indicates that the Dirac physics generates out-of plane spin polarization at the edges of the sample. Within the effective model, this can be explained by an anomalous velocity (compare discussion in Sections 4.2 and 9.4.1)

$$\mathbf{v}_{an} = \frac{1}{i}[\hat{\mathbf{r}}, \hat{H}_D] \propto \sigma_z \left(\mathbf{e}_z \times \nabla V\right) \tag{10.8}$$

which shifts spins $\uparrow,\downarrow$ into opposite directions and in the direction transverse to the potential, which can be the confinement potential but also an applied potential for the electrical bias. Therefore, this effect can be interpreted as the spin-Hall effect leading to the σ_z-polarization shown in Figure 10.3b, which is particularly large at the edges of the left and right leads. In the normal regime ($M > 0$), σ_z-polarization can be tuned by the change of the Fermi energy of the device or a Rashba coupling in the beam splitter part, however it is usually weaker than the in-plane spin polarization.

Further, in the signal $n_{hx}(\mathbf{r})$ in Figure 10.3a, we see that the helicity polarization generated by the SO barrier - which corresponds to an in-plane (σ_y) spin polarization of the effective 2-band model - is suppressed by precession of the spin around the effective $\mathbf{k}$-dependent magnetic field $\mathbf{B}_{\text{eff}} \propto \nabla V \times \mathbf{k}$. This precession is also visible as sign changes of n_{hx}, giving a blue/red pattern along the top right sample edge. The phase of this precession depends on E_F and therefore, good helicity polarization is obtained only for certain Fermi energies. This can be seen in Figures 10.5a and 10.6.

Formally, we can get rid of the valence band by setting $A = 0$, thus obtaining a 2-band model (2DEG) for only the conduction band, with j_{hx}/j identical to the normalized spin-y current. This 2-band model should not be confused with the effective 2-band model discussed above. In Figure 10.5a, the helicity current for the model with $A = 0$ (dotted black) shows oscillations due to wave interference, while different subband quantization leads to the opening of the lowest propagating mode for lower E_F than for the 4-band model. The helicity current for the 2-band model can be now compared with the helicity current of the full 4-band model. In the 4-band model, precession of the already polarized beam about $\mathbf{B}_{\text{eff}}$ leads to an additional structure of oscillation in comparison with 2-band model, so that the peaks in the helicity current become more isolated in the 4-band model and the signal becomes enhanced in comparison with the 2-band model.

For completeness, we have also analysed the helicity polarization in the parameter regime with band inversion ($M < 0$), where the band structure becomes topologically non-trivial. In the inverted regime, the effective model (10.7) does not account for the topologically protected edge states and therefore is not valid in general. However qualitatively, we still expect to see a competition between Rashba and Dirac physics. In particular, the edge states are polarized in the spin z-component when Rashba SO coupling is zero, and we expect them to partially suppress the h_x-polarization. In Figures 10.4, we show the normalized local helicity polarization $n_{hx}(\mathbf{r})/n(\mathbf{r})$ and the normalized σ_z-polarization $n_{\sigma z}(\mathbf{r})/n(\mathbf{r})$, this time for the inverted regime. The Rashba coupling constant is chosen by the condition $t_{SO}/t_0 = 0.1$, corresponding to $R = 75$ meV nm, and is non-zero only in the tilted barrier (shown in gray). We choose the Fermi energy $E_F = -17.6$ meV in the (electron-like) valence band with bulk gap in the range of $[-10, 10]$ meV. With these parameters, there are actually four propagating modes in the left/right leads. However, two first modes are spin edge states which, although the energy is outside of the bulk gap, do not merge with the bulk, and still retain their character of being strongly localized at the sample edges. They are almost fully polarized in σ_z (a deviation from full polarization is due

to overlap of edge states at the opposite sides of the lead). In the $n_{\sigma z}$ plot, they can be seen to bend, following the sample edges. Therefore, within our choice of the bias voltages, the charge signal T_{43} at the right lead is zero when only edge states are propagating. However, the next two propagating modes in the left lead pass the SO barrier from left to right, giving a non-zero T_{43} at E_F. Their interplay between the bulk and the edge states also contributes to T_{43}, as can been seen by analysing the S-matrix. The local σ_z-polarization in the right lead can be significant and is often higher than the in-plane (h_x) polarization. The lead-averaged helicity current for the inverted regime is in general smaller than for the normal regime. We do not show the polarization signals in the conduction band for the inverted regime. For small Fermi energies above the gap, heavy hole components are dominant, and consequently our analysis in the Section 10.6 shows that in-plane spin polarization will be small since it depends on $|E\pm\rangle$ components only. Although a relation between both polarizations is a complicated function of many parameters, we find that by changing E_F or normal versus inverted regimes, one can tune the ratio between in-plane and out-of plane polarizations.

10.4.4 Validity of the effective model

We make some rough estimates to show that for our parameters, we are in the right regime to consider the Dirac term of the effective model as small perturbation. We want to show that the expectation value of $\hat{H}_D = \frac{A^2}{4M^2}\left(\nabla V \times k\right)_z \sigma_z$ is small compared to $E_F = 13.8$ meV, which is the sweet spot. We may say that approximately, $|\langle \nabla V \times \mathbf{k}\rangle| < |\langle \nabla V\rangle| k_F$. The confinement potential should be of the order of E_F. If V is a step function with constant value zero inside, and the value E_F outside of the sample region, we can approximately say that $|\langle \nabla V\rangle| < E_F/W$, where W is a length scale corresponding to the sample width, which enters due to the normalization of the wave function. For W we enter the lead width 49 nm as lower bound, and we find that $\frac{Ak_F}{2M} \approx 0.5$ and $|\langle \hat{H}_D\rangle| < 0.1\ E_F$ for the sweet spot, confirming the validity of our effective model.

10.5 Detection scheme

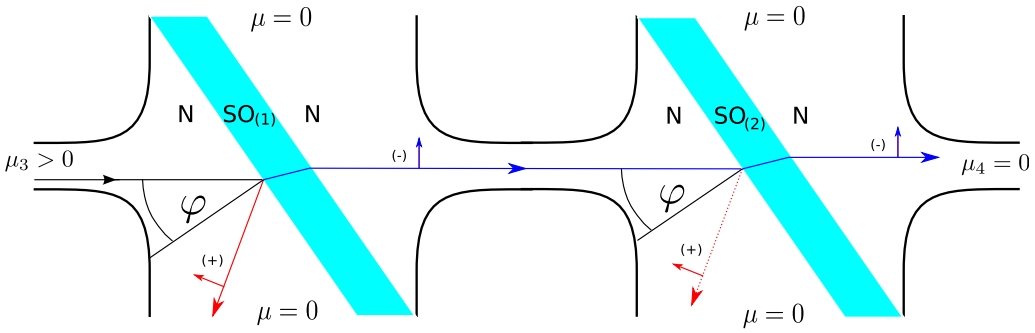
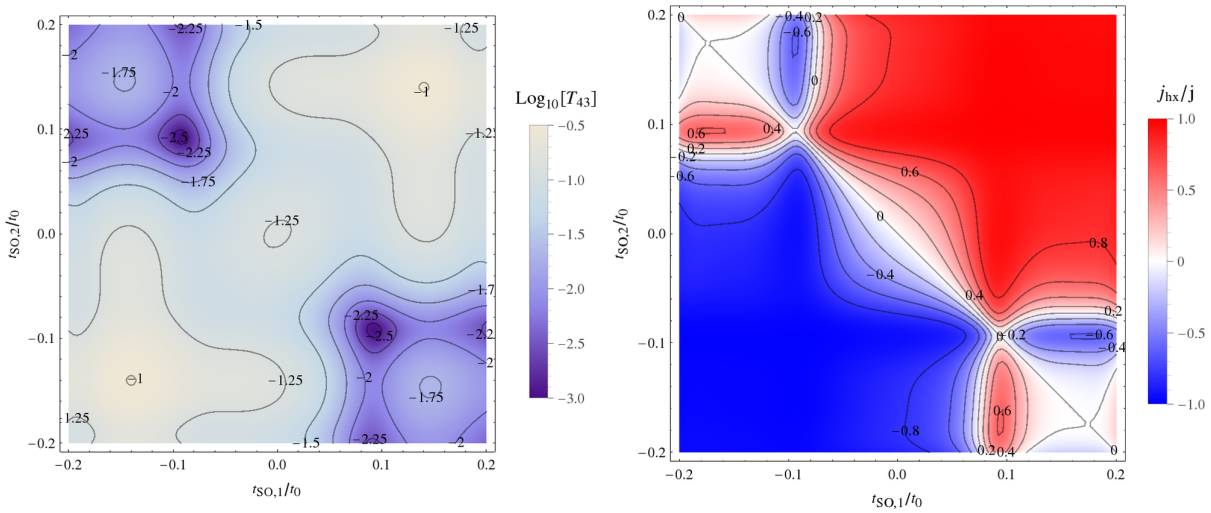


Figure 10.7: A double beam splitter setup with polariser (left) and analyser (right) can be used to detect the helicity current all-electrically. Taken from [Rot+14]. © (2014) by the American Physical Society

Experimentally, spin current (or even helicity current) detection is not established so far. Therefore, in this section, we show that by combining two beam splitter devices with indepen-



(a) (b)

Figure 10.8: The horizontal axis parametrizes the polariser (first device, $t_{SO,1}$) while the vertical axis parametrizes the analyser (second device, $t_{SO,2}$). (a) shows the transmission through the combined devices (from left to right), which is measurable all-electrically by transport. Here we tuned E_F to the maximal helicity current which corresponds to the maxima at $t_{SO,1} = t_{SO,2} = 0.1t_0$ and the minima at $t_{SO,1} = t_{SO,2} = -0.1t_0$. (b) shows the normalized helicity current in the right lead. Taken from [Rot+14]. © (2014) by the American Physical Society

dently tunable Rashba SO parameters in the barriers, an all-electrical detection of the helicity current is possible. As shown in Figure 10.7, the first device acts as polariser. The polarized beam then enters the second device. The tilting angle in both devices is assumed to be the same ($\varphi = 65°$). If the Rashba SO parameters are tuned to the same sign, the polarized beam will pass. With opposite signs, it can be blocked at the second device by total reflection. Thus, the transmission through the combined device, which is electrically measurable, can be used to prove that the current in the connecting part is polarized.

Instead of discretizing both devices on a common lattice, we can model a double beam-splitter device by combining the S-matrices of single beam splitters, $S^{(i)} = \begin{pmatrix} r^{(i)} & t'^{(i)} \\ t^{(i)} & r'^{(i)} \end{pmatrix}$ $(i = 1, 2)$. In our case (two propagating modes in the connecting leads), $r'^{(1)}$ and $r'^{(2)}$ are 2×2 matrices. The transmission amplitude through the combined device is [Dat07]

$$t = t^{(2)}(1 - r'^{(1)}r'^{(2)})^{-1}t^{(1)}. \tag{10.9}$$

Note that we have to be careful with the phase definitions of the S-matrices of device 1 and 2. Normally, the complex phases of the S-matrix entries are undefined and thus unrelated, since the mode basis consists of asymptotic states (i.e. they are evaluated far from the scattering region). However here, we need to use the same phase convention for an outgoing mode in device 1 as for the corresponding ingoing mode in device 2 and vice versa. Otherwise the result of (10.9) will be undefined. If the contribution of bound states in the intermediate lead becomes negligible, this method of combining S-matrices will be exact. This means the connecting lead should extend over several Fermi wave lengths. A geometric interpretation of t in terms of scattering paths can be obtained by expanding in the geometric series. The combined S-matrix can be used to

find helicity currents (see Appendix D).

Fig. 10.8a shows the transmission through a combination of polariser and detector as in the scheme of Fig. 10.7. The SO parameters of the barriers ($t_{SO,1}$ versus $t_{SO,2}$) can be tuned by two independent top gates. $t_{SO,i} = \frac{R_i}{2a}$ is the energy scale of Rashba SO in the discretized Hamiltonian. The horizontal axis parametrizes the polariser (first device, $t_{SO,1}$) and the vertical axis parametrizes the analyser (second device, $t_{SO,2}$).

The ratio of currents for parallel versus anti-parallel splitters, at $t_{SO,1} = 0.1t_0 = \pm t_{SO,2}$ is $0.066/0.0037 = 18$. Of course, if we wanted to find such a high ratio for other values of $t_{SO,i}$, we would need to detune also E_F. So experimentally, it would be desirable to have both a top and a bottom gate so that the Rashba parameter, controlled by the QW asymmetry, and E_F could be tuned independently. Fig. 10.8b shows the corresponding helicity current, normalized by the particle current. At $t_{SO,i} = 0.1t_0$, we find the value 0.997 (versus 0.0037 for anti-parallel splitters). Plot 10.8a is perfectly, and 10.8b nearly symmetric under exchange $t_{SO,1} \leftrightarrow t_{SO,2}$.

On the other hand, the detection of σ_z-polarization could be done by optical means like Faraday or Kerr rotation.

10.6 Relation between spin and helicity

So far, we have analysed polarization and transport in terms of the helicity operator, which we defined for this purpose. In this section, we want to address the obvious question, how much this observable has to do with physical (i.e. electron particle) spin polarization or currents.

For this purpose, we construct a local density matrix for the physical spin, starting from the envelope function $\psi(\mathbf{r})$:

$$\rho^s_{i,j}(\mathbf{r}) = \mathrm{Tr}\left[\tilde{\psi}(\mathbf{r})\tilde{\psi}^\dagger(\mathbf{r})|j\rangle_s\langle i|_s\right] \tag{10.10}$$

Here, $|i\rangle_s = \{|\uparrow\rangle_s, |\downarrow\rangle_s\}$ are basis functions of the physical electron spin. For the envelope function, we have changed the notation from $\psi(\mathbf{r})$ to $\tilde{\psi}(\mathbf{r})$, emphasizing that, while $\psi(\mathbf{r})$ is a 4-component vector depending on the in-plane coordinates $\mathbf{r} = (x, y)$, $\tilde{\psi}(\mathbf{r})$ should be understood as tensor product of in-plane and z-dependent factors and orbital and spin basis functions. The trace in (10.10) includes an integration over z. The in-plane envelope function will be expanded in components, $\tilde{\psi}(\mathbf{r}) = \sum_{i=1}^4 \psi_i(\mathbf{r})|i\rangle_{4b}$. Each of the $\mathbf{r}$-independent basis functions $|i\rangle_{4b} = \{|E+\rangle, |H+\rangle, |E-\rangle, |H-\rangle\}$ can again be expanded as $|i\rangle_{4b} = \sum_{j=1}^6 f_{i,j}(z)|j\rangle_K$ in terms of new envelope function components $f_{i,j}(z)$ that depend on the QW growth coordinate z, and Kane basis functions $|j\rangle_K = \{|\Gamma_6, \frac{1}{2}\rangle, |\Gamma_6, -\frac{1}{2}\rangle, |\Gamma_8, \frac{3}{2}\rangle, |\Gamma_8, \frac{1}{2}\rangle, |\Gamma_8, -\frac{1}{2}\rangle, |\Gamma_8, -\frac{3}{2}\rangle\}$. In order to finally have a basis suitable for evaluation of (10.10), the Kane basis functions are in turn expanded in the orbital and spin part $|j\rangle_K = \sum_{l=S,X,Y,Z}\sum_{s=\uparrow,\downarrow} c^j_{l,s}|l\rangle_o|s\rangle_s$.

We introduce the convention that Pauli matrices s_i act on the physical spin space. Note that σ Pauli matrices act on $+/-$ (Kramers') space and τ Pauli matrices on the E/H (QW subband) space. For convenience, we also define the matrices $s_\pm = \frac{s_x \pm i s_y}{2}$ and $s_{\uparrow/\downarrow} = \frac{1\pm s_z}{2}$.

As our result, we find that we can construct the space-dependent 2×2 density matrix $\rho^s(\mathbf{r})$ for the physical spin, starting from the 4-band wave function $\psi(\mathbf{r})$ and an $\mathbf{r}$-independent matrix,

$$\rho^s(\mathbf{r}) = \psi^\dagger(\mathbf{r}) \begin{pmatrix} \beta s_z + s_\downarrow & 0 & \beta s_- & 0 \\ 0 & s_\uparrow & 0 & 0 \\ \beta s_+ & 0 & -\beta s_z + s_\uparrow & 0 \\ 0 & 0 & 0 & s_\downarrow \end{pmatrix} \psi(\mathbf{r}). \tag{10.11}$$

Using the band parameters of Chapter 7.1.1, we find $\beta = 0.853$. Next, we make use of (10.11) to find the observables in the basis of the 4-band model, that represent physical spin components,

$$\mathrm{Tr}[s_{x/y}\rho^s(\mathbf{r})] = \beta\psi^\dagger(\mathbf{r})\sigma_{x/y}\tau_\uparrow\psi(\mathbf{r}) \tag{10.12}$$

with $\tau_\uparrow = \frac{\tau_0+\tau_z}{2}$. So if we compare local expectation values of $s_{y/x}$ and $h_{x/y}$, the difference is, that for the former, the heavy-hole components of the 4-band wave function do not contribute. This also means that generating in-plane spin polarization is not possible when transport is dominated by heavy holes.

We have calculated 2D density plots for $\sigma_y\tau_\uparrow$ in the same way as for h_x. Since for E_F in the conduction band (normal regime), the heavy-hole wave components are small, these plots (not shown) look mostly like the h_x-plots, with about 10% less efficiency in creating polarization (not taking into account the factor β). Also, spin currents can be defined in terms of $\sigma_y\tau_\uparrow$, but we prefer the observable h_x as measure of polarization since it is related to a conserved quantity.

For the spin-z component,

$$\mathrm{Tr}[s_z\rho^s(\mathbf{r})] = \psi^\dagger(\mathbf{r})\sigma_z(2\beta\tau_\uparrow - \tau_z)\psi(\mathbf{r}). \tag{10.13}$$

For $\beta = 1$, above observable reduces to $\sigma_z\tau_0$. We prefer the latter one as the measure of polarization since $[\hat{H}_N, \sigma_z\tau_0] = 0$.

10.7 Summary and outlook

In this chapter, we analysed a beam splitter device based on 2D topological insulators. We found that the Dirac-like model describing these materials can lead to higher in-plane or helicity polarization than the standard model utilizing only Rashba spin-orbit interaction. Further, in these systems in-plane and out-of plane spin polarization can be achieved. While the trivial insulator regime ensures strong in-plane polarization, in the topologically non-trivial regime the interplay between edge states and bulk states induces strong out-of plane polarization near the band gap. Several important relations between spin polarization and conserved quantities like helicity polarization are established as well as a simple all-electrical measurement scheme for in-plane spin current using two beam splitters is proposed. Although we focused on the parameters typical for HgTe quantum wells, the analysis presented here is also applicable to other systems described by the Hamiltonian of topological insulators, among them InAs/GaSb QWs [Liu+08] and Bi_2Se_3 thin films [Liu+10b]. We believe that tuning of the spin polarization from the out-of plane into the in-plane could be performed in one device based on InAs/GaSb QWs with top and bottom gates which change the positions of the electron and heavy-hole bands.

Chapter 11

Conclusion

The field of spintronics, which is concerned with information processing by use of the electron spin in addition to its charge, is mostly motivated by the prospect to allow for higher technological integration and higher switching speeds, than will be possible by charge-based information processing. In order to realize this vision, still a lot of scientific progress needs to be seen. Spin manipulation and spin transport are the main principles that need to be implemented. The main focus of this thesis was to analyse semiconductor systems where high fidelity in these principles can be achieved. To this end, we mainly used numerical methods for precise results, supplemented by simpler analytical models for interpretation. In this chapter we give a concluding presentation of the findings of this thesis, and briefly discuss some recent related scientific progress.

In Part I of this thesis, we introduced methods for construction of effective semiconductor heterostructure models, and for calculation of dynamical properties like transport. In Chapter 2, we outlined the derivation of the envelope function method after Burt, which is used to find effective multiband Hamiltonians for heterostructures. While earlier derivations of the envelope function method relied on the assumption of slowly varying perturbations of the $\mathbf{k}\cdot\mathbf{p}$ crystal Hamiltonian, so that their validity remains unclear in the interesting case of semiconductor heterostructures, the theory after Burt is based on the exact Schrödinger equation of the heterostructure and a series of successive controlled approximations, and therefore gives a better understanding of the validity of the envelope function approximation (EFA). Any quantitative result in this thesis is based on the method. In a wave matching technique applied to a problem including cubic Rashba spin-orbit terms, we show how technical difficulties with spurious solutions, boundary conditions and correct symmetrization of terms, can be eliminated. Since these issues often appear in the application of the EFA, the detailed description of our wave matching technique (Appendix F) might be useful for researchers working on similar problems.

The $\mathbf{k}\cdot\mathbf{p}$ and envelope function methods are very powerful when combined with the method of invariants under the symmetry of the underlying crystal. Chapter 3 contains an introduction to the method of invariants. We applied it to the geometric analysis of quadrupole Hamiltonians, which were needed in Chapter 8, where we demonstrated universal control of a single spin by the geometric concept of holonomy. The latter is also known as generalized Berry phase and is generated by adiabatic motion in parameter space. The underlying adiabatic theorem was presented in Chapter 4. We also applied the adiabatic theorem to derive the semiclassical theory of electronic transport, which is used for interpretation of numerical transport results. This method is closely related to the effective theory by Foldy for the non-relativistic limit of

the Dirac equation. Further, the topological invariants of the quantum Hall and quantum spin Hall states can be formulated in terms of the geometric notion of Berry curvatures.

Numerical transport calculations were the most important tool in this thesis. We presented the basis of transport theory in Chapter 5, with particular emphasis on calculation of transport observables in the Landauer-Büttiker formalism. While the formalism naturally works with spin-dependent or spin-orbit coupled Hamiltonians, there can be issues with the measurement of spin currents, if these are not conserved. We solved the problem by considering the lead-averaged spin current signal, an approach which turns out to be equivalent to a definition of spin current proposed by Shi [Shi+06] (compare Appendix E).

Chapter 6 contains a short introduction to topological states of matter and in particular, the 2D topological insulator (TI) system of HgTe/CdTe quantum wells (QW), which has attracted a lot of interest due to the experimental observation of the quantum spin Hall effect (QSHE). This material system is also interesting because it allows to realize ballistic transport of high quality, and because its Rashba spin-orbit (SO) coupling and electron density are tunable by electrical gating.

In part II of this thesis, my results were presented [Rot+10; Bud+12a; Rot+12; Rot+14]. In Chapter 7 we derived an effective four-band model including Rashba SO terms due to an applied potential that breaks the spatial inversion symmetry of the QW. It is an extension of the Bernvevig-Hughes-Zhang (BHZ) model for 2D TIs [Ber+06a]. We also demonstrated the consistency with the method of invariants. Further, we analysed spin transport in a reduced effective model for only the electron band, which already shows interesting physics because Rashba SO and the intrinsic (Dirac-like) SO of the BHZ model compete. The regime of validity of the reduced model is limited and its use is motivated by the simpler interpretation of numerical results. The spin precession due to Rashba SO is still visible in presence of Dirac-like SO terms and its amplitude can be increased. On the other hand, Rashba SO-induced spin precession reduces the spin polarization generated by the Dirac-like SO terms.

In Chapter 8 we showed that universal control of a single spin in a heavy-hole (HH) quantum dot is experimentally realizable without breaking time reversal invariance (TRI), but using a quadrupole field which is adiabatically changed as control knob. The mathematical control principle is a holonomy in the Kramers degenerate eigenspace attached at each point of the quadrupole parameter manifold. The holonomy can be visualized as generalization of a defect angle after parallel transport of a vector (belonging to the eigenspace), along a closed loop in the parameter manifold. The rotation operation of the spin takes the place of the defect angle. For experimental realization, we propose a GaAs/GaAlAs QW system hosting a lateral quantum dot. In this system, undesirable splitting of the HH/LH quadruplet in the ground state already appears without a quadrupole potential, because of spatial confinement. We propose to apply mechanical strain in order to compensate for this. Residual electrostatic perturbations will always be present in the experimental setup. We analysed their influence on the holonomy operation and showed their effect to be harmless. Together with all-electric spin pumping and spin filtering techniques [Bro+10b], and two-qubit computation gates by virtue of electrostatic gates [Los+98], our proposal provides a completion of the framework for TRI preserving quantum computation.

In Chapter 9 we analysed thermoelectric transport in four-terminal setups of HgTe/CdTe quantum wells. Due to spin-orbit coupling, an applied temperature gradient generates a transverse spin current. By analogy to the thermoelectric Nernst effect in presence of a perpendicular magnetic field, we call the signal spin Nernst effect. This effect is also closely related to the spin

Hall effect which was already analysed in Chapter 7.3, however in an effective two-band model instead of the four-band model. The spin Nernst effect and the spin Hall effect are found to be related by a Mott-like relation. We were able to qualitatively understand the spin Hall and spin Nernst signals in the metallic regimes by simple analytical models. For the parameter regime corresponding to an inverted band structure and for energies in the bulk gap of the QW, a peak in the spin Nernst signal is visible. It originates from the overlap of edge states at opposite sample edges, which is a finite size effect. Therefore, the spin Nernst coefficient could be used as experimental tool to analyse the finite-size induced mini-gap in the edge state spectrum.

We also analysed spin transport in the four-band model for HgTe/CdTe QWs in Chapter 10, but there, we included a Rashba SO coupling term corresponding to the result of Chapter 7. We proposed a beam splitter setup for all-electrical generation and detection of spin currents. Its working principle is similar to optical birefringence, and the setup is comparable to a proposal by Khodas [Kho+04], however we use a realistic finite sample geometry, and our material system shows more complicated physics because of the competition of different spin-orbit terms in the four-band model. We showed how demanding high efficiency in spin filtering predetermines the devices shape and symmetry. We analysed spin current and spin polarization signals of different spin vector components and showed that large in-plane spin polarization of the current can be obtained. The interplay of spin-orbit terms can even increase the amount of in-plane spin polarization for special parameters. The topologically trivial regime should be chosen to find high in-plane spin polarization, since edge states will be detrimental to this. On the other hand, the topologically non-trivial regime generally shows higher out-of plane spin polarization. With respect to tuning the direction of spin polarization, an important point is that heavy-hole components of the envelope function contribute only to out-of plane spin polarization, while electron-like components (they include an admixture of light holes) can contribute also to in-plane spin polarization. Locally, there can be significant out-of plane polarization of the spin.

Since spin is not a conserved quantity in the four-band model, we first analysed the transport of helicity, which is conserved quantum number in the four-band model even in presence of Rashba SO. Interestingly, the helicity operator (Appendix C) is not dependent on the band parameters, which is a fingerprint of the high symmetry of the effective model. We established the connection of helicity polarization to in-plane polarization of the physical spin. A similar approach is used for the out-of plane polarization of spin, which is also analysed indirectly. The definition of spin currents in systems with SO coupling has raised some discussion in the literature [Shi+06]. In our beam splitter setup, Rashba SO coupling is not present in the parts where spin currents are to be measured. This is an advantage for measurement of out-of plane spin currents. However, the device naturally generates in-plane spin polarization, and for this, the intrinsic SO coupling of the BHZ model introduces difficulties in the definition of spin currents. It is not possible to measure the in-plane polarized current in the most common and simple way, by introducing separate leads for spin basis states. We circumvented this problem by demonstrating a method that obtains the lead-averaged in-plane polarized spin current by combining scattering matrix entries with expectation values of helicity (Appendix D).

The analyses and methods shown in this thesis can be applied to many related scientific questions. Although we focused on the parameters typical for HgTe quantum wells, our analysis of spin transport is also applicable to other systems described by the Hamiltonian of topological insulators, among them InAs/GaSb QWs [Liu+08] and Bi_2Se_3 thin films [Liu+10b].

The numerical methods used in this thesis can also be applied to the analysis of transport in presence of a magnetic field. Sensitivity of the topological protection of the QSH state

to a magnetic field that breaks time reversal symmetry is an interesting and relevant topic. Experimental measurements [Kön+07] have shown that the transport of the QSHE edge states breaks down already at moderate fields, showing larger sensitivity to out-of plane than to in-plane magnetic fields. Topological protection of charge transport will break down if in addition to the presence of a magnetic field, coupling of opposite spins is possible (e.g. due to structural or bulk inversion asymmetry-induced SO), and if backscattering is caused by disorder. In this context, bulk inversion asymmetry (BIA)-induced SO is especially interesting since it is present even in symmetrically grown QWs, and since its effect on the edge state dispersion is more important than Rashba SO for small momenta. The values for BIA can be calculated in a similar manner as Rashba (SIA) SO, employing $\mathbf{k}\cdot\mathbf{p}$ theory supplemented by the methods of invariants. The leading term is small compared to the bulk gap [Kön+08]. A numerical analysis of the combined influence of magnetic field, BIA and disorder on the QSHE [Mac+10] reproduces the experimentally observed cusp-like suppression of the edge states conductance. A simple picture for understanding this breakdown in terms of scattering matrices for pairs of edge states has been given in [Del+12]. The idea is to describe scattering between nearby pairs of edge states by a time reversal symmetric S-matrix, and combine this with S-matrices for edge states enclosing some magnetic flux (magnetic flux impurities), where Aharonov-Bohm phases enter. While low magnetic fields are sufficient to destroy the QSHE, in high magnetic fields, a transition to the quantum Hall effect must happen. This transition has been analysed in [Tka+10] and relies on backscattering that locally breaks time reversal invariance, like the before mentioned magnetic flux impurities.

In the context of magnetic field combined with spin-orbit coupling, another interesting effect is the crossover from weak localization, i.e. enhanced backscattering due to constructive interference of paths related by time reversal symmetry in a system without SO coupling, to weak antilocalization. The latter is the fingerprint of the different behaviour of a SO-coupled system under time reversal and can be explained in terms of a Berry phase. Experimentally, weak antilocalization has recently been observed in thin wires fabricated of HgTe/CdTe QWs, in the normal and inverted regimes [Müh+14a].

To close, let us briefly take an outlook on related scientific progress in the field. The QSHE opens new possibilities for spintronics devices based on edge states. One proposal is a spin transistor [Kru+11] which benefits from the intrinsic BIA-induced SO interaction in HgTe/CdTe quantum wells and allows for gate-controlled spin or charge switching between edge channels. Further, the edge states of the QSHE were combined with the SHE, to demonstrate spin polarization of the edge states [Brü+12]. With regard to the generation of spin currents for spintronic applications, an open question is how one could possibly benefit from the dissipationless spin current, which naturally appears in the helical Dirac cone of the metallic surface states of a 3D TI.

Even if manipulation of the spin is not the goal, the topologically protected charge transport in a TRI system opens many new possibilities, among them proposed efficient thermoelectric systems, which could pave the way for future heat to power conversion [Tre+10].

In this thesis, the concept of holonomic quantum computation [Zan+99] was only applied to the manipulation of a single spin. However, the general concept allows for operations on larger Hilbert spaces, and has attracted a lot of interest because of the prospect to implement fault tolerant quantum computation, which is topologically protected in the sense that generated unitary operations on the Hilbert space will depend only on the topology of a braiding group [Nay+08]. For such topological quantum computation, quasiparticles showing non-Abelian statistics are

required. An interesting candidate are Majorana fermions, which could be realized experimentally in hybrid systems consisting of topological insulator wires interfaced with a superconductor [Fu+08]. Despite a lot of experimental work, convincing experimental proof of a Majorana state is still outstanding.

Concerning the area of topological insulators, certainly one direction of further research is the theoretical prediction and experimental study of new material systems. Several systems have been experimentally confirmed to be 3D topological insulators [And13], and an InAs/GaSb/AlSb heterostructure system has been both predicted to be a 2D topological insulator [Liu+08] and experimentally confirmed, by proving the existence of helical edge states is this system [Kne+11]. Further research on material systems is needed to improve possibilities for technological integration of TIs, in particular hybrid structures of TIs with superconductors and ferromagnets will be of interest. For practical applications, systems with a large bulk gap will be required, in order to finally allow device operation at room temperature.

On the theoretical side, an interesting subject is extending the scope of topological classification. Time reversal invariance, which is the protecting symmetry of the topological insulator phase, has been supplemented by the discrete anti-unitary symmetry of charge conjugation, in order to find a complete periodic table of topological insulators and topological superconductors in different spatial dimensions [Sch+08]. The topological classification of phases has been further extended to include the point group of the underlying crystal structure as protecting symmetry. In a situation without spin-orbit coupling but with time reversal invariance, the new class of topological crystal insulators has been predicted [Fu11].

With regard to the topological classification of correlated, i.e. strongly interacting phases, it will be interesting to explore interaction driven quantum phase transitions, with respect to the topology of the occurring phases. An intriguing subject is how phase classification by conventional order parameters, which capture the physics of broken symmetries, can be intertwined with topological classification of phases. An example is the prediction of a topological Mott insulator [Rag+08], a phase that is insulating due to strong Coulomb interaction, but hosts topologically protected edge states inside the insulating bulk gap. In this sense, the system shows an interaction driven, topological phase transition. A continuous change of a conventional order parameter indicates the phase transition, which is at the same time accompanied by a discontinuous jump of the topological invariant. Recently, another example of intertwined physics of thermodynamic and topological phase transitions has been elucidated by the analysis of a Hubbard-like interacting model based on the BHZ model [Ama+14]. In this system, it has been shown that for sufficiently strong Coulomb interaction, the topological phase transition between band insulator and topological insulator happens without closing the bulk gap, and that it becomes a thermodynamic phase transition, which is associated with experimentally accessible fingerprints.

Appendix A

Kubo formula for conductivity from linear response

In this appendix, outline the derivation of the Kubo formula [Kub57] for the conductivity from linear response theory. The linear response formalism allows to evaluate non-equilibrium, i.e. irreversible processes like currents and dissipation, if they are small, so that they can be treated as perturbation to the equilibrium. The time dependent expectation value of an observable is thus given by the equilibrium expectation value of a correlation function that describes the observables fluctuation in time. There are many different derivations and equivalent expressions for the Kubo formula. Here, our goal is to obtain the Kubo formula in the form that is is used as starting point in Kohmotos [Koh85] discussion of the quantum Hall effect.

We assume that the dynamics of a system is given by a time-independent Hamiltonian H_0 plus a small perturbation $H'(t) = \hat{B}F(t)$, where $\hat{B}$ is an operator, and $F(t)$ is real with $F(-\infty) = 0$. We are interested in the time dependent expectation value of an operator $\hat{O}$,

$$\langle \hat{O} \rangle(t) = \mathrm{Tr}(\rho(t)\hat{O}), \tag{A.1}$$

where the density matrix obeys the von Neumann equation $\frac{d}{dt}\rho(t) = i[\rho(t), H_0 + H'(t)]$, and we use the ansatz $\rho(t) = \rho_0 + \hat{f}(t)$. Here $\rho_0 = \frac{1}{Z}e^{-\beta H_0}$ is the time-independent equilibrium density matrix, with $Z = \mathrm{Tr}(\rho_0)$. In the interaction picture, the dynamics of $\hat{f}_I(t) = e^{iH_0 t}\hat{f}(t)e^{-iH_0 t}$ is found to be

$$\frac{d}{dt}\hat{f}_I(t) = -i[H'_I(t), \rho_0] + \mathcal{O}(H'^2). \tag{A.2}$$

The time-dependent expectation value up to first order in $H'(t)$ is found by integration,

$$\langle \hat{O} \rangle(t) = \langle \hat{O} \rangle_0 - i\int_{-\infty}^{t} dt' \,\mathrm{Tr}\left(\hat{O}_I(t)[H'_I(t'), \rho_0]\right) \tag{A.3}$$

$$= \langle \hat{O} \rangle_0 - i\int_{0}^{\infty} dx \,\langle [\hat{O}_I(t), H'_I(t-x)] \rangle_0 \tag{A.4}$$

where $\langle ... \rangle_0$ denotes the statistical expectation value evaluated with ρ_0.

Since we are interested in the conductivity tensor, we will apply this to find current expectation values as function of a driving electric field, which is taken as perturbation. For this, one can either implement the driving field by the dipole energy term $H'(t) = e\sum_i \mathbf{r}_i \mathbf{E}(t)$, where the

$\mathbf{r}_i$ are the positions of the electrons and $\mathbf{E}(t)$ is the electric field. This is how Kubo proceeds in his original work [Kub57]. But we choose a different approach, which must lead to equivalent results because it is related to the dipole energy by a gauge transformation, and implement the driving field in the vector potential. In order to distinguish vector potential of the driving electric field and a possible constant magnetic field, we write $\mathbf{A}(t) = \mathbf{A}_B + \mathbf{A}_E(t)$, with $\mathbf{A}_E(t)$ chosen such that $\mathbf{E}(t) = -\frac{\partial}{\partial t}\mathbf{A}(t) = \mathbf{E}(\omega)e^{-i(\omega+i\eta)t}$. The infinitesimal $\eta > 0$ is used to model a slow turning on of the perturbation. For $\hat{O}$, we insert the current component μ

$$J_\mu = -e\frac{\partial H}{\partial k_\mu} = J_\mu^{(p)} + J_\mu^{(d)} \tag{A.5}$$

where H is the full Hamiltonian, including the perturbation. Therefore, the current contains a term independent of the perturbing $\mathbf{A}_E$, which is called paramagnetic current $J_\mu^{(p)}$, as well as a term linear in $\mathbf{A}_E$, which is called diamagnetic current $J_\mu^{(d)}$.

In the expansion (A.4), the term $\langle J_\mu\rangle_0$ gives a contribution of the diamagnetic current, while the paramagnetic current of a parity-symmetric system must vanish in equilibrium. The diamagnetic contribution is parallel to the direction of the applied field, $\langle J_\mu\rangle_0 = K_\mu A_{E,\mu}(t)$ with $K_\mu = -e\langle\frac{\partial^2 H}{\partial k_\mu^2}\rangle_0$. Being interested in the response linear in $\mathbf{A}_E$, we only need the paramagnetic current part for the commutator term. Note that the perturbation $H'(t) = H|_{\mathbf{A}} - H|_{\mathbf{A}_E=0}$ contains a part proportional to A_E^2, which was important to obtain the diamagnetic current. However, in the commutator expression in (A.4), we drop the part that depends on A_E^2 and insert $H'(t) \approx -\mathbf{J}^{(p)} \cdot \mathbf{A}_E(t)$. Then, the conductivity tensor $\sigma_{\mu\nu}(\omega)$ is found from the defining relation

$$\langle J_\mu(\omega)\rangle = \frac{1}{2\pi}\int dt\, e^{i\omega t}\langle J_\mu\rangle(t) = \sigma_{\mu\nu}(\omega)E_\nu(\omega). \tag{A.6}$$

Since the time dependent factor $\mathbf{A}_E(t)$ in $H'(t)$ is not an operator, it may be pulled out of the commutator expectation value in (A.4). Then, one can use $[H_0, \rho_0] = 0$ to show that the latter no longer depends on t. We find

$$\sigma_{\mu\nu}(\omega) = \lim_{\eta\to 0}\frac{1}{i(\omega+i\eta)}\left(K_\mu\delta_{\mu\nu} + i\int_0^\infty dt\, e^{i(\omega+i\eta)t}\langle\left[J_\mu^{(p)}(0), J_\nu^{(p)}(-t)\right]\rangle\right). \tag{A.7}$$

In the following, we consider the static Hall conductivity $\sigma_{xy}(0)$ for a 2DEG on a lattice, with a non-zero perpendicular magnetic field. As discussed in [Koh85], we need to restrict ourselves to magnetic fields corresponding to a rational number of flux quanta per lattice site, in order to find a basis of generalized Bloch states. The momenta $\mathbf{k}$ used in the following should be understood as elements of the magnetic Brillouin zone, and the indices n, m will be used to label the band of the generalized Bloch states. The eigenstate representation of the equilibrium density matrix in the grand canonical ensemble is diagonal $\mathbf{k}$ and given by $\rho_0(\mathbf{k}) = \sum_m f(E_m(\mathbf{k}))P_m(\mathbf{k})$, with the projectors $\{P_m(\mathbf{k})\}_m$ on the generalized Bloch states, and the Fermi distribution function $f(E_m)$ at eigenenergy $E_m(\mathbf{k})$. In this basis, the Hamiltonian is given by $H_0(\mathbf{k}) = \sum_n E_n(\mathbf{k})P_n(\mathbf{k})$. Note that we need only the current-current correlation of the paramagnetic current parts, which can be already found from $H_0(\mathbf{k})$ without the perturbation. After evaluating the time integral in (A.7), one arrives at

$$\sigma_{xy}(\omega) = \lim_{\eta\to 0}\frac{i}{\omega+i\eta}\sum_{\mathbf{k},m,n}\left(f(E_m(\mathbf{k})) - f(E_n(\mathbf{k}))\right)\frac{X_{nm}}{\omega+i\eta+E_m(\mathbf{k})-E_n(\mathbf{k})}, \tag{A.8}$$

where we have defined current-current correlation matrix elements

$$X_{nm} = \mathrm{Tr}(J_x^{(p)}(\mathbf{k})P_n(\mathbf{k})J_y^{(p)}(\mathbf{k})P_m(\mathbf{k})), \tag{A.9}$$

for which $X_{mn} = X_{nm}^*$. Let us drop the parameter $\mathbf{k}$ for convenience and introduce $\Delta_{mn} = E_m(\mathbf{k}) - E_n(\mathbf{k})$. We can rewrite (A.8) by explicitly adding up summands with $m \leftrightarrow n$ interchanged, and restricting the sum to the ordering $E_m(\mathbf{k}) < E_n(\mathbf{k})$. The restricted sum is denoted as $\sum'$. Further, at zero temperature, the $f(E_m)$ will take only values 0 or 1, which gives the restriction that the Fermi energy must lie in a gap, $E_m(\mathbf{k}) < E_F < E_n(\mathbf{k})$. We split σ_{xy} into two parts

$$\sigma_{xy}(\omega) = \sigma_{xy}^{(I)}(\omega) + \sigma_{xy}^{(II)}(\omega), \tag{A.10}$$

$$\sigma_{xy}^{(I)}(\omega) = \lim_{\eta\to 0} {\sum_{\mathbf{k},m,n}}' (f(E_m) - f(E_n)) \frac{i(X_{nm} - X_{nm}^*)}{(\omega + i\eta)^2 - \Delta_{mn}^2} \tag{A.11}$$

$$\sigma_{xy}^{(II)}(\omega) = \lim_{\eta\to 0} i {\sum_{\mathbf{k},m,n}}' (f(E_m) - f(E_n)) \frac{(X_{nm} + X_{nm}^*)\Delta_{mn}}{((\omega + i\eta)^2 - \Delta_{mn}^2)(\omega + i\eta)}. \tag{A.12}$$

We are especially interested in the DC limit $\omega \to 0$. Luckily, ω and $i\eta$ only appear as sum, so we do not need to specify the order of taking their limits. Without the small η, we wouldn't be able to work within the equilibrium distribution given by ρ_0. $\sigma_{xy}^{(I)}(\omega)$ is the important part, which gives the non-zero (quantum) Hall conductance, and it is the starting point in [Koh85]. We can use symmetry under rotation about $\hat{z}$ by $\frac{\pi}{2}$ to show that the part $\sigma_{xy}^{(II)}(\omega)$ vanishes. To see this, we note that the rotational symmetry gives $\sigma_{y,-x}(\omega) = \sigma_{xy}(\omega) = -\sigma_{yx}(\omega)$. The only factor in $\sigma_{xy}(\omega)$ that depends on current directions is X_{nm}. By interchanging current coordinates, one finds $X_{nm}(J_x^{(p)} \leftrightarrow J_y^{(p)}) = X_{mn}$. Making use of this property and demanding invariance of the analytic expression for $\sigma_{xy} = \sigma_{y,-x}$, we find a condition that the matrix elements X_{nm} must fulfil,

$$X_{nm} + X_{nm}^* \stackrel{!}{=} -X_{mn} - X_{mn}^* = -(X_{nm}^* + X_{nm}) = 0, \tag{A.13}$$

and thereby, we have $\sigma_{xy}^{(II)}(\omega) = 0$ by symmetry. At zero temperature and in the DC limit, the Hall conductance simplifies to

$$\sigma_{xy} = \frac{1}{i} \sum_{\substack{E_n(\mathbf{k}) < E_F < E_m(\mathbf{k}) \\ \mathbf{k},n,m}} \frac{(J_y^{(p)})_{nm}(J_x^{(p)})_{mn} - (J_x^{(p)})_{nm}(J_y^{(p)})_{mn}}{(E_n - E_m)^2} \tag{A.14}$$

Analogue to (A.11) and (A.12), one may define the longitudinal conductivity parts $\sigma_{xx}^{(I)}$ and $\sigma_{xx}^{(II)}$. The part $\sigma_{xx}^{(I)}$ will vanish since X_{nm} becomes real. In the case of no magnetic field and in the zero temperature limit, $\sigma_{xx}^{(I)}$ will also vanish, leaving only the diamagnetic part, which diverges for the DC limit $\omega \to 0$. In order to obtain a meaningful (non-diverging) result for the DC limit, one needs to include a scattering mechanism like disorder or phonons [Czy07]. However, this problem does not affect the Hall conductivity or longitudinal conductivity for $\omega \neq 0$. In the calculation shown here, we need to assume a perfect lattice so we can work with the basis of generalized Bloch states.

Appendix B

Quasi-degenerate perturbation theory

Quasi-degenerate perturbation theory, also called Löwdin partitioning [Low51], is a formulation of time-independent perturbation theory that is particularly useful for the application in $\mathbf{k} \cdot \mathbf{p}$ theory, where an infinite-dimensional eigensystem must be reduced to an effective theory written in the basis of the finite set of the most important bands. The transformation is also discussed by Foldy [Fol+50] in his seminal work on a non-relativistic effective theory for a spin-$\frac{1}{2}$ particle in a perturbing electromagnetic field.

The method can be seen as a generalization of the standard time-independent perturbation theory [Sak04]. Instead of corrections to the eigenenergies, one obtains corrections to a submatrix of the Hamiltonian, which account for the coupling to other blocks of the Hamiltonian. In the special case where the submatrix is just a number, the two formulations become identical. In the standard perturbation theory, special care must be taken if the energy level of interest is degenerate or nearly degenerate. One needs to use a basis, which will first diagonalize the perturbation in the degenerate set, before developing the perturbation series. The theory discussed here avoids this in a natural way, since it operates with the full submatrix of the (quasi-) degenerate set.

The derivation given here follows closely the one given in Appendix B of Winkler [Win05]. Let the set A denote the (quasi-)degenerate set of states or bands, and the set B all other states or bands. This defines a partitioning of the Hamiltonian H into blocks. The goal is to find an (approximate) unitary transformation

$$\tilde{H} = e^{-S} H e^{S} \tag{B.1}$$

so that $\tilde{H}$ is block-diagonal. S must be anti-hermitian, $S^\dagger = -S$. Further, we choose S as block-off-diagonal by ansatz.

We split the Hamiltonian into parts

$$H = H^0 + H' = H^0 + H^1 + H^2 \tag{B.2}$$

where the part H^0 is already diagonal, with energies $\{E_n\}_n$ on the diagonal, and the perturbation H' is assumed to be small compared to the energy gap between states of set A (for which we will use indices m, m', m'') and states of set B (which will be denotes with indices l, l', l''). The perturbation is further split into a block-diagonal part H^1, i.e. $H^1_{ml} = H^1_{lm} = 0$, and a block-off-diagonal part H^2, i.e. $H^2_{mm'} = H^2_{ll'} = 0$.

The transformed Hamiltonian can be written as

$$\tilde{H} = \sum_{j=0}^{\infty} \frac{1}{j!} [H, S]^{(j)}, \tag{B.3}$$

where

$$[A, B]^{(j)} = [...[[A, \underbrace{B], B], .., B}_{\text{j times}}]. \tag{B.4}$$

(B.3) can be proven order by order. For the contribution of order j in the perturbation, we can first ignore non-commutation of S and H and start from the term $\frac{1}{j!}(-S+S)^j H$. We expand the power and restore the correct ordering of terms according to (B.1) by shifting all powers of $-S$ to the left of H, and all powers of $+S$ to the right of H. The result is equal to $\frac{1}{j!}[H, S]^{(j)}$. A very useful observation is that for a product involving n block-off-diagonal matrices, the result is block-off-diagonal for n odd and block-diagonal for n even. With this knowledge, we can split $\tilde{H}$ into a block-diagonal part $\tilde{H}_d$, and a block-off-diagonal part $\tilde{H}_n$,

$$\tilde{H}_d = \sum_{j=0}^{\infty} \frac{1}{(2j)!} [H^0 + H^1, S]^{(2j)} + \sum_{j=0}^{\infty} \frac{1}{(2j+1)!} [H^2, S]^{(2j+1)}, \tag{B.5}$$

$$\tilde{H}_n = \sum_{j=0}^{\infty} \frac{1}{(2j+1)!} [H^0 + H^1, S]^{(2j+1)} + \sum_{j=0}^{\infty} \frac{1}{(2j)!} [H^2, S]^{(2j)}. \tag{B.6}$$

Now, S is found by the condition that the off-diagonals should vanish, $\tilde{H}_n = 0$. To this end, we write $S = S^{(1)} + S^{(2)} + ...$ and successively find the corrections $S^{(j)}$ of order j in the perturbation,

$$[H^0, S^{(1)}] = -H^2 \tag{B.7}$$

$$[H^0, S^{(2)}] = -[H^1, S^{(1)}] \tag{B.8}$$

$$[H^0, S^{(3)}] = -[H^1, S^{(2)}] - \frac{1}{3}[[H^2, S^{(1)}], S^{(1)}] \tag{B.9}$$

$$[H^0, S^{(4)}] = -[H^1, S^{(3)}] - \frac{1}{3}[[H^2, S^{(1)}], S^{(2)}] - \frac{1}{3}[[H^2, S^{(2)}], S^{(1)}] \tag{B.10}$$

$$\begin{aligned}[H^0, S^{(5)}] = -\frac{1}{45} \Big(& 45\,[H^1, S^{(4)}] + 15\,[[H^2, S^{(1)}], S^{(3)}] + 15\,[[H^2, S^{(2)}], S^{(2)}] \\ & + 15\,[[H^2, S^{(3)}], S^{(1)}] - [[[[H^2, S^{(1)}], S^{(1)}], S^{(1)}], S^{(1)}] \Big)\end{aligned} \tag{B.11}$$

Since $S^{(j)}$ is block-off-diagonal, we can solve for all non-zero matrix elements, $S^{(j)}_{ml} = \frac{1}{E_m - E_l}[H^0, S^{(j)}]_{ml}$. The energy denominator never vanishes, since by assumption there is an energy gap between the sets A and B.

Entering the expression for S in (B.5) and sorting by orders in the perturbation, one finds $\tilde{H}$, which is block-diagonal. Including the perturbation up to third order, we have

$$\begin{aligned}\tilde{H}_{mm'} = & H^0_{mm'} + H'_{mm'} + \frac{1}{2} \sum_l H'_{ml} H'_{lm'} \left(\frac{1}{E_m - E_l} + \frac{1}{E_{m'} - E_l} \right) \\ & - \frac{1}{2} \sum_{l,m''} \left(\frac{H'_{ml} H'_{lm''} H'_{m''m'}}{(E_{m'} - E_l)(E_{m''} - E_l)} + \frac{H'_{mm''} H'_{m''l} H'_{lm'}}{(E_m - E_l)(E_{m''} - E_l)} \right) \\ & + \frac{1}{2} \sum_{l,l'} H'_{ml} H'_{ll'} H'_{l'm'} \left(\frac{1}{(E_m - E_l)(E_m - E_{l'})} + \frac{1}{(E_{m'} - E_l)(E_{m'} - E_{l'})} \right)\end{aligned} \tag{B.12}$$

In principle, this scheme can also be used to carry out the block-diagonalization numerically up to any order, i.e. find the block-diagonalization up to numerical machine precision. But this would only be useful if for some reason, one wants to restrict the unitary transformation to those generated by block-off-diagonal matrices. Otherwise, one should simply perform a complete numerical diagonalization. However, we will only make use of the transformation in the analytical form (B.12).

It is also important to note that (B.12) remains valid if the entries of the matrix H' are operators. In this case, it is important to keep the order in which matrix elements appear in (B.12).

Appendix C

Helicity operator in 4-band model[1]

C.1 Requirements for $h(\mathbf{k})$

In order to define a helicity operator $\hat{h}$ in the 4-band model (10.1), we state some requirements that match the usual definition of the helicity operator, $\frac{\sigma\cdot\mathbf{k}}{k}$, for a spin-1/2 particle. These requirements will make our definition unique up to an overall sign. $\hat{h}$ should be

- Hermitian,
- time reversal symmetric: $[\mathcal{T}, \hat{h}] = 0$.

  For the time reversal operator $\mathcal{T} = ZK$ with Z unitary and K being the complex conjugation, we use the convention $Z = -i\sigma_y\tau_0$, where σ Pauli matrices act on $+,-$ space and τ Pauli matrices act on E, H space, and τ_0 is a unit matrix. Further, $\mathcal{T}\hat{\mathbf{k}}\mathcal{T}^{-1} = -\hat{\mathbf{k}}$ and $\mathcal{T}\mathbf{k}\mathcal{T}^{-1} = \mathbf{k}$, because $\hat{\mathbf{k}} = -i\nabla$ is an operator and $\mathbf{k}$ a real vector.
- parity odd: $P\hat{h}P^\dagger = -\hat{h}$, with $P\hat{\mathbf{k}}P^\dagger = -\hat{\mathbf{k}}$,
  $P|E\pm\rangle = -|E\pm\rangle$ and $P|H\pm\rangle = |H\pm\rangle$,
- $\hat{h}$ should have only eigenvalues ± 1,
- and finally, we demand $[\hat{H}, \hat{h}] = 0$.

Note that $[\hat{H}, \hat{h}] = 0$ implies that $\hat{h}$ will also have the rotational and translational symmetry of $\hat{H}$, where the latter implies that we can write $\hat{h} = \int d^3k\, |\mathbf{k}\rangle h(\mathbf{k}) \langle\mathbf{k}|$. In the following it will be shown that $h(\mathbf{k})$ takes the form of Eq. (10.3). The rotational symmetry is about the direction $\hat{z}$, which is the growth direction of the QW, and is given by $D_\alpha h(\phi) D_{-\alpha} = h(\phi + \alpha)$, where $\phi = \arg(k_+)$ and $D_\alpha = \exp(-iS_z\alpha)$ and $S_z = \mathrm{diag}(\frac{1}{2}, \frac{3}{2}, -\frac{1}{2}, -\frac{3}{2})$.

We may say that parity-oddness of $\hat{h}$ is its defining feature, because it corresponds to the parity-oddness of the spin-orbit terms. We call branches of a dispersion related by time reversal, if their crossing at $k = 0$ is enforced by Kramers' degeneracy. The basic idea for the definition of $h(\mathbf{k})$ is to assign different signs ± 1 to states at the same $\mathbf{k}$, if they are lying on dispersion branches related by time reversal symmetry. Kramers partners will be assigned the same helicity. If $\lambda_\pm(\mathbf{k})$ are the eigenvalues of $h(\mathbf{k})$, this is exactly what the combination of parity-oddness and time reversal symmetry ensures, because $\lambda_+(\mathbf{k}) \overset{P}{=} -\lambda_+(-\mathbf{k}) \overset{\mathcal{T}}{=} -\lambda_-(\mathbf{k})$. Since SO coupling

[1] This appendix has also been published in [Rot+14].

removes the degeneracy of bands related by TRS, the observable $\hat{h}$ is suitable to detect SO-related effects.

C.2 Symmetry based derivation of $h(\mathbf{k})$

We perform a construction of the helicity operator by symmetry, similar to the construction of $H(\mathbf{k})$ in Chapter 7.1.2.

The point group T_d of the zinc blende structure of HgTe is reduced to the C_{nv} symmetry group by the quantum well confinement (n depending on the direction of growth). In the axial approximation that we used for derivation of $H(\mathbf{k})$, we had the point group $C_{\infty v}$, and time reversal symmetry.

The point group includes reflections at a plane including $\hat{z}$, e.g. the element $\hat{C}_v$ with $\hat{C}_v(k_x, k_y)\hat{C}_v^{-1} = (-k_x, k_y)$ and $\hat{C}_v|E\pm\rangle = |E\mp\rangle$, $\hat{C}_v|H\pm\rangle = |H\mp\rangle$. If we demand invariance of $H(\mathbf{k})$ under this symmetry, this enforces the parameters A, R, S and T to be real. Note that this result relies on the conventions we use for $\mathcal{T}$ and $\hat{C}_v$.

We look at the decomposition of a 4x4 matrix in terms of the Clifford algebra. In Chapter 7.1.2, we have constructed the most general diagonal-in-k, rotational invariant about $\hat{z}$, time-reversal symmetric and parity-odd Hamiltonian in the basis of S_z-eigenstates. The Hamiltonian was the part H_{SO} of (10.1), which depends on the Rashba parameters R, S, T. We introduce new parameters r, s, t that take the places of kR, k^2S, k^3T, and obtain the most general ansatz

$$h(\mathbf{k}) = \begin{pmatrix} 0 & 0 & -irk_-/k & -isk_-^2/k^2 \\ 0 & 0 & isk_-^2/k^2 & itk_-^3/k^3 \\ ir^*k_+/k & -is^*k_+^2/k^2 & 0 & 0 \\ is^*k_+^2/k^2 & -it^*k_+^3/k^3 & 0 & 0 \end{pmatrix} \tag{C.1}$$

Because of the reflection symmetry C_v of our system, R, S and T are real. However, it is not yet clear that r, s and t will be real. Evaluating $[h(\mathbf{k}), H(\mathbf{k})] = 0$, we obtain

$$R\,\mathrm{Im}[r] + kS\,\mathrm{Im}[s] = 0 \tag{C.2}$$
$$S\,\mathrm{Im}[s] + kT\,\mathrm{Im}[t] = 0 \tag{C.3}$$
$$-Rs - kSt + kSr^* + k^2Ts^* = 0 \tag{C.4}$$
$$2\mathcal{M}s + Ak(r - t) = 0 \tag{C.5}$$

If we say that these equations must be true independent of the parameters R, S and T and for all k, we obtain $s = 0$ and $r = t$ and real. If we set $r = t = 1$ in order to fix the eigenvalues to ± 1, we are done. With this convention, the conduction band states with the higher energy at the same k, are assigned positive helicity (assuming $R, S, T > 0$).

With a bit more calculation, we even do not need to assume the solution to be independent of R, S and T, and we still obtain the same, unique result. Let us assume $R, T \neq 0$, then

$$\mathrm{Im}[r - t] = \left(\frac{S}{kT} - \frac{kS}{R}\right)\mathrm{Im}[s] = -\frac{2\mathcal{M}}{Ak}\mathrm{Im}[s]$$

Except for very special parameters of $H(\mathbf{k})$, we may conclude $\mathrm{Im}[s] = 0$, and thus $\mathrm{Im}[r] = 0$ and $\mathrm{Im}[t] = 0$. Now that we know that s is real, we may use (C.4) and (C.5),

$$s = \frac{kS}{k^2T - R}(t - r) = \frac{Ak}{2\mathcal{M}}(t - r)$$

so we must have $t = r$ except for very special parameters, and $s = 0$. Fixing eigenvalues to ± 1, we obtain the result (10.3).

C.3 Projector based derivation of $h(\mathbf{k})$

From the derivation by symmetry, it was not yet very clear, that the different signs of helicity correspond to branches of the band structure that are related by time reversal symmetry. To clarify this, we give here a derivation based on projectors on eigenstates of $H(\mathbf{k})$.

For an eigenstate $|\psi_{\mathbf{k}}\rangle = |\mathbf{k}\rangle|\chi_+(\mathbf{k})\rangle$ with spinor $|\chi_+(\mathbf{k})\rangle$, we use the time reversal operator to define a related spinor $|\chi_-(\mathbf{k})\rangle$ at the same $\mathbf{k}$:

$$\mathcal{T}|\psi_{-\mathbf{k}}\rangle = |\mathbf{k}\rangle Z|\chi_+(-\mathbf{k})\rangle^* =: |\mathbf{k}\rangle|\chi_-(\mathbf{k})\rangle. \tag{C.6}$$

Since we need the state at $-\mathbf{k}$ to find the related spinor, it is important that we have a set of eigenvectors that are continuous functions of $\mathbf{k}$. It is not possible to find eigenvectors as continuous functions in the complete plane of (k_x, k_y), but for any given direction of $\mathbf{k}$, e.g. $k_y = 0$, it is possible to find eigenvectors that are continuous on this line.

We make use of the spectral representation,

$$H(\mathbf{k}) = \sum_{\alpha=E,H} \sum_{s=+,-} |\chi_{\alpha,s}(\mathbf{k})\rangle\langle\chi_{\alpha,s}(\mathbf{k})| E_{\alpha,s}(\mathbf{k}). \tag{C.7}$$

Due to the degeneracy at $\mathbf{k} = 0$, we have to be careful how to define eigenfunctions $|\chi_\pm(\mathbf{k})\rangle$ as functions of $\mathbf{k}$, such that they are continuous functions. We set $E_{\alpha,\pm}(-k) = E_{\alpha,\mp}(k)$. Then, $E_{\alpha,\pm}(k)$ are differentiable functions even at the band crossing at $k = 0$. We may define $|\chi_{\alpha,-}(k)\rangle = \mathcal{T}|\chi_{\alpha,+}(-k)\rangle$. States are continuous functions on cuts $\phi = \arg(k_+) = const. + n\pi$ where n takes values 0 and 1.

Then we can rewrite $h(\mathbf{k})$ using projectors on states, simply by replacing the eigenenergies with ± 1.

$$h(\mathbf{k}) = \sum_{\alpha=E,H} s_\alpha \mathrm{sgn}(k_x) \left(P_{\alpha,+}(\mathbf{k}) - P_{\alpha,-}(\mathbf{k})\right) \tag{C.8}$$

with $P_{\alpha,\pm}(\mathbf{k}) = |\chi_{\alpha,\pm}(\mathbf{k})\rangle\langle\chi_{\alpha,\pm}(\mathbf{k})|$ and signs $s_\alpha = \pm$ which are to be determined. By construction, $[H(\mathbf{k}), h(\mathbf{k})] = 0$. Under time reversal $H(\mathbf{k}) \to ZH^*(-\mathbf{k})Z^\dagger$, we have $P_{\alpha,\pm}(\mathbf{k}) \to P_{\alpha,\mp}(\mathbf{k})$ so that $[\hat{h}, \mathcal{T}] = 0$.

If we assume that at $\mathbf{k}$ the bands are non-degenerate, this construction of $h(\mathbf{k})$ is unique up to the signs s_α, and we have to show that $s_E = s_H = 1$ coincides with our earlier definition. The question of signs is equivalent to the question how we would like to label (e.g. numerically obtained) eigenstates with indices $+$ and $-$. In particular, the relative sign s_E/s_H is relevant. The reasonable choice is, that when considering the limit $R, S, T \to 0$, the $+$ eigenvectors should be continuously connected to eigenvectors of the upper left 2×2 block of $H(\mathbf{k})$. Instead of considering a continuous deformation of the parameter space of the Hamiltonian, we will use here a particle-hole symmetry to obtain an equivalent result. As we will show in the next section, even though $H(\mathbf{k})$ is not particle-hole symmetric, there is an operation P_{eh} (see Eq. (C.10)) with

$$P_{eh}|\chi_{E,\pm}\rangle\langle\chi_{E,\pm}|P_{eh}^\dagger = |\chi_{H,\pm}\rangle\langle\chi_{H,\pm}|. \tag{C.9}$$

P_{eh} should not be confused with the parity operator P. From (C.8) it is clear that with $s_E = s_H$, we find $[h(\mathbf{k}), P_{eh}] = 0$, while this does not hold for $s_E = -s_H$. Since in the last section, the construction by symmetry gave a unique (parity-odd) result, for which it can be checked that $[h(\mathbf{k}), P_{eh}] = 0$, we must have $s_E = s_H$ in order to fulfil parity-oddness of $h(\mathbf{k})$. A more direct proof of the parity-oddness of (C.8) would be desirable, but is not easy since a single projector $P_{\alpha,+}(\mathbf{k})$ does not have definite parity.

C.4 Particle-hole symmetry of states

The spectrum of $H(\mathbf{k})$ is not particle-hole symmetric, both because the spin-orbit terms for the E and H bands differently depend on $\mathbf{k}$, and because of $\epsilon(k)$. However, there is a particle-hole symmetry of the eigenstates of $H(\mathbf{k})$. For the discussion here, we set $\epsilon(k) = 0$ without loss of generality. We postulate the operator

$$P_{eh}(\phi) = \begin{pmatrix} & k_+/k & & \\ -k_-/k & & & \\ & & & -k_-/k \\ & & k_+/k & \end{pmatrix} \tag{C.10}$$

where $\phi = \arg(k_+) = \arg(k_x + ik_y)$ and $k = |\mathbf{k}|$. We will show that P_{eh} has the meaning of a particle-hole symmetry, in the sense of (C.9). Obviously, P_{eh} does not depend on k, and it has the properties $P_{eh}^\dagger = -P_{eh} = P_{eh}^{-1}$. Being interested in $H(\mathbf{k})$ as a function of the parameters R and T, we denote this with an index. We find the relation

$$-P_{eh} H_{RT} P_{eh}^\dagger = H_{-Tk^{-2},-Rk^2}. \tag{C.11}$$

Thereby, we recognize that for $R = T = 0$, the operator P_{eh} gives the particle-hole symmetry $\{P_{eh}, \hat{H}\} = 0$. Further, $[P_{eh}, \hat{h}] = 0$, as is easily checked using the representation (10.3) in the main text. Therefore, if $|\chi\rangle$ is an helicity eigenstate, $P_{eh}|\chi\rangle$ is also an eigenstate of the same helicity. Now we want to prove that if $|\chi\rangle$ is an energy eigenstate, $P_{eh}|\chi\rangle$ is also an energy eigenstate, which is not yet clear from (C.11) for the case $R, T \neq 0$. Thus we can find the complete set of eigenvectors just starting with two states $|\chi_\pm\rangle$ having helicity ± 1.

In the following, we use the notation $|\chi_{n,\mathbf{k}}\rangle = |\chi_{n,\phi}\rangle$ for energy eigenvectors, when dependency on k is unimportant. TRS relates states of $\mathbf{k}$ to states of $-\mathbf{k}$. Since we are looking for a relation between states of the same $\mathbf{k}$, we use a combination of time reversal and a rotation by π. The rotational symmetry is $D_{-\phi} H(\phi) D_\phi = H(0)$.

Since energies at $k \neq 0$ are non-degenerate, we know that time reversal will map spinors onto themselves[2], $Z|\chi^*_{n,-\mathbf{k}}\rangle = ZD^*_\pi|\chi^*_{n,\mathbf{k}}\rangle = e^{i\phi_n}|\chi_{n,\mathbf{k}}\rangle$. The state-dependent phase $e^{i\phi_n}$ is unimportant because it can be removed by re-definition. Let us define $U := (ZD^*_\pi)^\dagger = i\sigma_x\tau_z$. Then we can replace the operation of complex conjugation by the unitary operation U, $|\chi^*_n\rangle = U|\chi_n\rangle$. Since $[P_{eh}(0), U] = [i\sigma_z\tau_y, i\sigma_x\tau_z] = 0$, we exploit the rotational symmetry $D_{-\phi} P_{eh}(\phi) D_\phi = P_{eh}(0)$ to evaluate the commutator at $\phi = 0$.

[2] Here we consider just the spinor part of the state. The full state $|\psi_\mathbf{k}\rangle = |\mathbf{k}\rangle|\chi_{n,\mathbf{k}}\rangle$ is of course mapped to it's Kramers' partner, which is attached at $|-\mathbf{k}\rangle$ and thus is orthogonal to $|\mathbf{k}\rangle$. Note also that using the rotation to go from $\mathbf{k} \to -\mathbf{k}$ corresponds to the continuous change of the eigenvector on the circle $k = \text{const.}$, so we end up with a state of same energy. If we did a continuous change of the eigenvector by keeping $\phi = \text{const.}$ as in the previous section, we would end up with a state of different energy.

Our goal is to prove

$$\langle\chi_{n,\phi}|P_{eh}(\phi)|\chi_{n,\phi}\rangle = 0. \tag{C.12}$$

Eigenstates of different energy, but same $\mathbf{k}$, must be orthogonal, and the SO terms remove degeneracy. Since the eigenspaces of $h(\mathbf{k})$ are just 2-dimensional, finding a state of same helicity that is orthogonal, suffices to show that it is an eigenstate of energy. Thus (C.12) is equivalent to (C.9). Since $P_{eh}^\dagger = -P_{eh}$, and $|\chi_{n,\phi}\rangle = D_\phi|\chi_{n,0}\rangle$, we have

$$\begin{aligned} &-\langle\chi_{n,\phi}|P_{eh}(\phi)|\chi_{n,\phi}\rangle = -\langle\chi_{n,0}|P_{eh}(0)|\chi_{n,0}\rangle \\ &= \langle\chi_{n,0}|P_{eh}(0)|\chi_{n,0}\rangle^* = \langle\chi_{n,0}|U^\dagger P_{eh}^*(0)U|\chi_{n,0}\rangle \\ &= \langle\chi_{n,0}|U^\dagger P_{eh}(0)U|\chi_{n,0}\rangle = \langle\chi_{n,0}|P_{eh}(0)|\chi_{n,0}\rangle = 0. \end{aligned} \tag{C.13}$$

C.5 Using $h(\mathbf{k})$ in the transport code

For a state $|\mathbf{k}\rangle(a|\chi_+(\mathbf{k})\rangle + b|\chi_-(\mathbf{k})\rangle)$ (which is not eigenstate of $\hat{H}$), we intend to define our spin transport by measuring $|a|^2 - |b|^2$. In transport calculations, one typically considers eigenstates of energy at the Fermi level. Then, the $\mathbf{k}$-vectors in general will be different:

$$|\psi\rangle = a|\mathbf{k}_1\rangle|\chi_+(\mathbf{k}_1)\rangle + b|\mathbf{k}_2\rangle|\chi_-(\mathbf{k}_2)\rangle \tag{C.14}$$

Here we include just two modes, because for typical parameters and a given direction of $\mathbf{k}$, $H(\mathbf{k})$ will have just two propagating modes at the Fermi level, and we expect evanescent modes to have negligible effect on transport. Because the Hamiltonian is rotationally invariant, the direction of propagation coincides with the direction of $\mathbf{k}$, and we assume $\mathbf{k}_1/k_1 = \mathbf{k}_2/k_2$. The helicity operator just depends on the direction $\mathbf{k}/k$. On the one hand, we have

$$\langle\psi|\hat{h}|\psi\rangle = \langle\psi|\left(a|\mathbf{k}_1\rangle h(\mathbf{k}_1)|\chi_+(\mathbf{k}_1)\rangle + b|\mathbf{k}_2\rangle h(\mathbf{k}_2)|\chi_-(\mathbf{k}_2)\rangle\right) = |a|^2 - |b|^2. \tag{C.15}$$

Instead of calculating this global expectation value, we may as well calculate the local expectation value $\langle\chi|h(\mathbf{k}_i)|\chi\rangle$ (i=1 or 2) using just the 4-component spinor $|\chi\rangle := a|\chi_+(\mathbf{k}_1)\rangle + b|\chi_-(\mathbf{k}_2)\rangle$, to obtain the same result,

$$\langle\chi|h(\mathbf{k}_i)|\chi\rangle = |a|^2 - |b|^2 + b^*a\langle\chi_-(\mathbf{k}_2)|\chi_+(\mathbf{k}_1)\rangle - a^*b\langle\chi_+(\mathbf{k}_1)|\chi_-(\mathbf{k}_2)\rangle = |a|^2 - |b|^2. \tag{C.16}$$

Here we used the projector representation of $h(\mathbf{k})$ and made use of the fact that the expectation value must be real. If we want to numerically evaluate the local expectation value of $h(\mathbf{k})$, we only need to choose a fixed direction of $\mathbf{k}$, and we can analyse the contribution of the modes for a wave going in this direction.

Appendix D

Obtaining bias-dependent observables from the S-matrix[1]

In Chapter 10, we have seen that the helicity, i.e. eigenvalue of $h(\mathbf{k})$ for a given direction of $\mathbf{k}$, corresponds to the components of the electron beam that will be split at a barrier where the Rashba SO is different from the background. Since the right lead of the considered finite-geometry device is constructed in a way to guide the beam in the direction $k = k_x$, we consider the operator $h_x = h(k = k_x) = \sigma_y \tau_z$, for which the expectation value is called local helicity density, and the associated current is called helicity current. Although $[H(\mathbf{k}), h(\mathbf{k})] = 0$, we have $[\hat{H}, h_x] \neq 0$ even with zero Rashba SO terms. Thus, the common practice to calculate spin currents, by introducing separate leads for the spin directions, does not work here.

Instead, we show a method for the helicity current calculation, which combines transmission coefficients of the S-matrix with operator expectation values, evaluated for propagating states in the leads. The advantage is that this method is even applicable with Rashba SO coupling in the leads. In that case, the local spin/helicity current oscillates as function of the position in the lead, but we will be interested in its average value only[2].

A lead is connected to a contact at one end and to the scattering region at the other end. Thanks to the reflectionless property of the contacts [Sza+89; Dat07], for a lead l, the ingoing modes will be populated with a Fermi distribution $f_0(E - \mu_l)$, while the outgoing modes will be populated by electrons that originate from ingoing modes of other leads and which pass the scattering region.

Now we construct a density matrix ρ for the ingoing states α, β. We take α as combined index $\alpha = (l_\alpha, k_\alpha, n_\alpha)$ of lead, momentum (x component, i.e. along the direction of the lead) and mode index. We use the short notations $\delta_{\alpha,\beta} = \delta_{l_\alpha, l_\beta} \delta(k_\alpha - k_\beta) \delta_{n_\alpha, n_\beta}$ and $\int \mathrm{d}\alpha = \sum_{l_\alpha} \int \mathrm{d}k_\alpha \sum_{n_\alpha}$. ϵ_α is the subband dispersion of lead l_α and $v_\alpha = \partial \epsilon_\alpha / \partial k_\alpha$ is the velocity of mode n_α.

$$\rho = \int \mathrm{d}\gamma \, f_0(\epsilon_\gamma - \mu_{l_\gamma}) \theta(v_\gamma) |\gamma\rangle\langle\gamma| \tag{D.1}$$

In order to obtain a density matrix ρ' for the outgoing states, we propagate the states with the

[1] This appendix has also been published in [Rot+14].

[2] In a lead oriented along x-direction and with non-zero Rashbas SO, the average spin-x and spin-z currents will cancel, but the spin-y current may be non-zero.

S-matrix like $|\gamma\rangle \to S|\gamma\rangle$,

$$\rho'_{\alpha,\beta} = \int \mathrm{d}\gamma\, f_0(\epsilon_\gamma - \mu_{l_\gamma}) \langle\alpha|S|\gamma\rangle\langle\gamma|S^\dagger|\beta\rangle. \tag{D.2}$$

We have to specify how to evaluate S-matrix elements

$$\begin{aligned}\langle\alpha|S|\gamma\rangle &= \langle k_\alpha, n_\alpha|S|k_\gamma, n_\gamma\rangle = \sqrt{|v_\alpha v_\gamma|}\,\theta(v_\gamma)\theta(-v_\alpha)\delta(\epsilon_\alpha - \epsilon_\gamma) t_{l_\alpha n_\alpha, l_\gamma n_\gamma} \\ &= \int d\mu' \frac{1}{\sqrt{|v_\alpha v_\gamma|}} \delta(k_\alpha - k_\alpha(\mu'))\delta(k_\gamma - k_\gamma(\mu')) t_{l_\alpha n_\alpha, l_\gamma, n_\gamma},\end{aligned} \tag{D.3}$$

where the current-normalized transmission amplitudes appear, i.e. the matrix of the $t_{l_\alpha n_\alpha, l_\gamma n_\gamma}$ is unitary, and describes elastic scattering. The velocity factors ensure current conservation. The Heaviside θ factors select only ingoing modes in lead l_γ and outgoing modes in lead l_α, and the set of $k_\alpha(\mu')$ are the outgoing solutions of $\epsilon_\alpha(k) = \mu'$ for the set of subbands n_α.

We are now prepared to evaluate a local observable $\hat{O}(x)$, with coordinate x lying in some lead p, relative to the lead's coordinate system. We assume that all terms in $\hat{O}(x)$ contain a factor P_x (projector on x). We need the matrix elements in the basis of propagating (outgoing) states. They can be written in the form

$$O_{\beta,\alpha} = \langle\beta|\hat{O}(x)|\alpha\rangle = \delta_{l_\alpha,p}\delta_{l_\beta,p}\langle\chi_\beta|O(x, k_\beta, k_\alpha)|\chi_\alpha\rangle e^{i(k_\alpha - k_\beta)x} \tag{D.4}$$

with some matrix-valued function $O(x, k_\beta, k_\alpha)$ and the normalized transverse modes $\{|\chi_\alpha\rangle\}$. For the (helicity or spin) current operators, $O(x, k_\beta, k_\alpha)$ does not depend on x, and for the local (helicity or spin) density operators, it also does not depend on k_α and k_β (see Section 10.4.2 for the definition of the used operators $\hat{O}(x)$). Then, the expectation value of $\hat{O}(x)$ depends on in- and outgoing modes,

$$\langle\hat{O}(x)\rangle = \mathrm{Tr}[(\rho + \rho')\hat{O}(x)]. \tag{D.5}$$

We leave equilibrium by introducing lead-dependent bias voltages $\mu_l = E_F + \delta\mu_l$. If we calculate the response at lead p due to some bias at lead $q \neq p$, ρ will not contribute.

$$\begin{aligned}\frac{\delta\langle\hat{O}(x)\rangle}{\delta\mu_q} &= \int \mathrm{d}\alpha \int \mathrm{d}\beta\, \frac{\partial\rho'_{\alpha,\beta}}{\partial\mu_q} O_{\beta,\alpha} = \int \mathrm{d}\alpha \int \mathrm{d}\beta \int \mathrm{d}\gamma\, \sqrt{|v_\alpha v_\beta|}|v_\gamma| t_{\alpha,\gamma} t^*_{\beta,\gamma} \delta(\epsilon_\alpha - \epsilon_\gamma)\cdot \\ &\quad \cdot \delta(\epsilon_\gamma - \epsilon_\beta)\theta(-v_\alpha)\theta(-v_\beta)\theta(v_\gamma)\frac{\partial f_0(\epsilon_\gamma - \mu_\gamma)}{\partial\mu_p} O_{\beta,\alpha}\end{aligned} \tag{D.6}$$

Let us assume zero temperature, so $\theta(v_\gamma)\frac{\partial f_0(\epsilon_\gamma - \mu_\gamma)}{\partial\mu_q} = \frac{1}{|v_\gamma|}\delta(k_\gamma - k^n_q)\delta_{l_\gamma,q}$, where k^n_q is the momentum of the ingoing mode n in lead q for energy μ_q.

In general, $O_{\beta,\alpha}$ will not be diagonal and therefore, it will also show oscillations in x. But if $O(x, k_\beta, k_\alpha)$ is x-independent and we are interested in a mean value, we can still simplify (D.6) using

$$\delta_{l_\alpha,l_\beta}\delta(\epsilon_\alpha - \epsilon_\beta)\int_0^L \mathrm{d}x\, e^{-i(k_\alpha - k_\beta)x} \to \frac{L}{|v_\alpha|}\delta_{\alpha,\beta} \tag{D.7}$$

for large L. Put into words, if the lead and momentum indices are the same, the band indices must also be the same if energies are the same. Note that, if we have degeneracy of energy and

momentum, this relation will not hold. For the lead-averaged response of the expectation value (e.g. for the conductance), we find

$$\frac{1}{L}\int_0^L \mathrm{d}x \frac{\delta\langle\hat{O}(x)\rangle}{\delta\mu_q} = \int \mathrm{d}\alpha \int \mathrm{d}\gamma\, |v_\gamma||t_{\alpha,\gamma}|^2 \delta(\epsilon_\alpha - \epsilon_\gamma)\theta(v_\gamma)\frac{\partial f_0(\epsilon_\gamma - \mu_\gamma)}{\partial \mu_p} O_{\alpha,\alpha}\delta_{l_\alpha,p} \tag{D.8}$$

$$\overset{T=0}{=} \int \mathrm{d}\alpha \sum_{n_\gamma} |t_{\alpha,\gamma}|^2 \delta(\epsilon_\alpha - \epsilon_\gamma) O_{\alpha,\alpha}\delta_{l_\alpha,p} \tag{D.9}$$

$$= \sum_{n_\alpha}\sum_{n_\gamma} |t_{\alpha,\gamma}|^2 \frac{1}{|v_\alpha|} O_{\alpha,\alpha} \tag{D.10}$$

where the solutions $k_\alpha(E_F)$ of lead q should be entered whenever integration over α is no longer present. The derivation shown here is mostly standard, apart from the averaging step (D.7). It is this step which keeps our result quite general and simple at the same time. In the literature, most of the time $\hat{O}(x)$ is taken as the current operator. Then, averaging is not necessary since $O_{\beta,\alpha}$ is already diagonal.

For the case of degeneracy in energy and momentum, the transverse mode basis used to evaluate $O_{\alpha,\alpha}$ will matter, although numerical diagonalisation will choose an arbitrary basis. In particular, this applies to the situation without Rashba SO in the leads, where subbands are degenerate. There are two ways to fix this problem: Firstly, we can generalize (D.10) to the case where some subbands are degenerate, by keeping track of the coherence of the degenerate subset,

$$\frac{1}{L}\int_0^L \mathrm{d}x \frac{\delta\langle\hat{O}(x)\rangle}{\delta\mu_q} \overset{T=0}{=} \sum_{m,n,l}{}' t_{m,l}t^*_{n,l}\frac{1}{|v_n|} O_{n,m} \tag{D.11}$$

where we replaced the collective Greek indices by the Latin mode indices, being the only ones of interest here. The sum is over all modes m,n in lead p and l in lead q, both at the Fermi energy. Since we are interested only in average over the lead, the summation over m,n is restricted to pairs with $k_m = k_n$, implying $v_n = v_m$. If there are only two propagating modes, and if we do not include Rashba SO terms in the lead, i.e. when all modes are degenerate, formula (D.11) will be also correct without the averaging over the x-coordinate (because the result is constant).

Alternatively, we may get rid of the degeneracy by adding a tiny perturbation which will fix the mode basis in the leads. The form of the perturbation will depend on the operator $\hat{O}(x)$. E.g, if we introduce a magnetic field B_z as perturbation, this will cause spin precession about z and therefore suppress the σ_y-polarization even in the limit $B_z \to 0$, thus changing the physics. In order to check that the small perturbation does not change the physics, we have to prove that the general result (D.11) reduces to (D.10) in the limit of the vanishing perturbation. It turns out, in the 2-band model (2DEG with Rashba SO), when interested in σ_y-polarization or currents, we may use either a small magnetic field B_y or a small Rashba SO coupling. In the 4-band model, things are more complicated. But we find that again, a small Rashba SO term does the job and does not change the physics.

Appendix E

Continuity equation and vanishing average torque[1]

We would like understand how the helicity current that we analysed in Chapter 10, is connected to the helicity polarization. Therefore, in this appendix we derive a generalized continuity equation. For a spin-$\frac{1}{2}$ system with Rashba SO, the continuity equation including a source term reads [Shi+06]

$$\frac{\partial S_l}{\partial t} + \nabla \cdot \mathbf{J}^{(l)} = \mathcal{T}_l \tag{E.1}$$

with the local spin density $S_l = \frac{1}{2}\psi^\dagger(\mathbf{r})\sigma_l\psi(\mathbf{r})$, local spin current $\mathbf{J}^{(l)} = \mathrm{Re}\left(\psi^\dagger(\mathbf{r})\frac{1}{4i}\{[\hat{\mathbf{r}}, H], \sigma_l\}\psi(\mathbf{r})\right)$ and the local spin source (torque) $\mathcal{T}_l = \mathrm{Re}\left(\psi^\dagger(\mathbf{r})\frac{1}{2i}[\sigma_l, H]\psi(\mathbf{r})\right)$, and we have put $\hbar = 1$.

We generalize this result to the 4-band model (10.1), which is an effective model for the 4-component envelope function $\psi(\mathbf{r})$. Since the quadratic and cubic terms of H_{SO} are unimportant (see main text), we consider only the linear Rashba terms. Our goal is to find an equation similar to (E.1), but the Pauli matrix σ_l should be replaced by a general 4×4 matrix Ξ with constant entries (no operators). To obtain helicity density, current, and torque terms, we may later choose $\Xi = h_x = \sigma_y\tau_z$. To find density, current and torque terms related to the spin-z polarization, we may choose $\Xi = \sigma_z\tau_0$. We start with a standard derivation of current matrix elements, which is also appropriate for finding our Ξ-currents.

We consider the general n-band model

$$H = H_0 + V(\hat{\mathbf{r}}) + U(\hat{\mathbf{r}}, \hat{\mathbf{k}}) \tag{E.2}$$

where $H_0 = \sum_{i,ml} D_{i,ml}\hat{k}_m\hat{k}_l|i\rangle\langle i|$ is a diagonal[2] matrix quadratic in $\hat{\mathbf{k}}$ and $D_{i,ml}$ is a band-dependent (inverse) effective mass tensor, $V(\hat{\mathbf{r}})$ is a general Hermitian $n \times n$ matrix representing a local potential, and U is linear in $\hat{\mathbf{k}}$ and Hermitian, of the form

$$U(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \sum_i \alpha_i(\hat{\mathbf{r}})\hat{k}_i\alpha_i(\hat{\mathbf{r}}) \tag{E.3}$$

with a vector $\boldsymbol{\alpha}(\hat{\mathbf{r}})$ of $n \times n$ matrices, here with 2 components. We note in passing that for another common symmetrization of non-commuting operators, $U_A(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \sum_i\{\alpha_i(\hat{\mathbf{r}}), \hat{k}_i\}$, the resulting continuity equation is of the same form.

[1] This appendix has also been published in [Rot+14].

[2] It should be possible to generalize the calculation to general non-diagonal matrices, if required.

For arbitrary wave functions ψ_n, ψ_m, we consider the overlap

$$w_{nm}(\mathbf{r}) = \psi_n^\dagger(\mathbf{r})\psi_m(\mathbf{r}) = \langle\psi_n|\mathbf{r}\rangle\langle\mathbf{r}|\psi_m\rangle. \tag{E.4}$$

We apply the Schrödinger equation to get

$$\partial_t w_{nm}(\mathbf{r}) = \langle\psi_n|\frac{1}{i}\Big[|\mathbf{r}\rangle\langle\mathbf{r}|, H\Big]|\psi_m\rangle. \tag{E.5}$$

Here, terms with $V(\hat{\mathbf{r}})$ cancel because they are local. As usual, we exploit that H_0 consists of second order derivatives. Pulling out a derivative ∇, after some steps we arrive at

$$\partial_t w_{nm}(\mathbf{r}) = -\frac{1}{2}\nabla\cdot\langle\psi_n|\Big\{|\mathbf{r}\rangle\langle\mathbf{r}|, \frac{1}{i}[\hat{\mathbf{r}}, H_0]\Big\}|\psi_m\rangle + \langle\psi_n|\frac{1}{i}\Big[|\mathbf{r}\rangle\langle\mathbf{r}|, U\Big]|\psi_m\rangle, \tag{E.6}$$

where the anticommutator $\{X, Y\} = XY + YX$ appears. For the U term, we use

$$\langle\psi_n|\frac{1}{i}\Big[|\mathbf{r}\rangle\langle\mathbf{r}|, U\Big]|\psi_m\rangle = -\sum_j \partial_j\langle\psi_n|\mathbf{r}\rangle\alpha_j^2\langle\mathbf{r}|\psi_m\rangle = -\frac{1}{2}\nabla\langle\psi_n|\Big\{|\mathbf{r}\rangle\langle\mathbf{r}|, \frac{1}{i}[\mathbf{r}, U]\Big\}|\psi_m\rangle. \tag{E.7}$$

Since $[\hat{\mathbf{r}}, H_0] = [\hat{\mathbf{r}}, H_0 + V(\hat{\mathbf{r}})]$, we obtain the continuity equation $\partial_t w_{nm} = -\nabla\cdot\mathbf{j}_{nm}$ with current matrix elements

$$\mathbf{j}_{nm}(\mathbf{r}) = \langle\psi_n|\frac{1}{2}\Big\{|\mathbf{r}\rangle\langle\mathbf{r}|, \boldsymbol{\mathcal{V}}\Big\}|\psi_m\rangle, \tag{E.8}$$

where $\mathcal{V}_j = \frac{\partial H}{\partial k_j} = \frac{1}{i}[\hat{r}_j, H]$. The diagonals are real and give the well-known current expression

$$\mathbf{j}_{nn}(\mathbf{r}) = \mathrm{Re}\left(\psi_n^\dagger(\mathbf{r})\frac{1}{i}[\hat{\mathbf{r}}, H]\psi_n(\mathbf{r})\right). \tag{E.9}$$

Based on this calculation, it is not difficult to find our generalized continuity equation for the current of Ξ (e.g. helicity),

$$\frac{\partial S_\Xi}{\partial t} + \nabla\cdot\mathbf{J}^{(\Xi)} = \mathcal{T}_\Xi \tag{E.10}$$

with the local Ξ-density $S_\Xi = \psi^\dagger(\mathbf{r})\Xi\psi(\mathbf{r})$, local Ξ-current $\mathbf{J}^{(\Xi)} = \mathrm{Re}\left(\psi^\dagger(\mathbf{r})\frac{1}{2i}\{[\hat{\mathbf{r}}, H], \Xi\}\psi(\mathbf{r})\right)$ and the local Ξ-source (torque) $\mathcal{T}_\Xi = \mathrm{Re}\left(\psi^\dagger(\mathbf{r})\frac{1}{i}[\Xi, H]\psi(\mathbf{r})\right)$.

We define the projector on coordinate $\mathbf{r}$, $P_\mathbf{r} = |\mathbf{r}\rangle\langle\mathbf{r}|$. With the Schrödinger equation and its Hermitian conjugate, we obtain, similar to (E.5),

$$\partial_t\langle\psi_n|\mathbf{r}\rangle\Xi\langle\mathbf{r}|\psi_m\rangle = \frac{1}{i}\langle\psi_n|[P_\mathbf{r}\Xi, H]|\psi_m\rangle. \tag{E.11}$$

The diagonal element ($m = n$) gives $\frac{\partial S_\Xi}{\partial t}$. Since $[P_\mathbf{r}, \Xi] = 0$, we have

$$\frac{1}{i}[P_\mathbf{r}\Xi, H] = \frac{1}{2i}\{[P_\mathbf{r}, H], \Xi\} + \frac{1}{2i}\{P_\mathbf{r}, [\Xi, H]\}. \tag{E.12}$$

The diagonal matrix element of the last term gives the Ξ-torque $\mathcal{T}_\Xi = \frac{1}{2i}\langle\psi_n|\{P_\mathbf{r}, [\Xi, H]\}|\psi_n\rangle$.

For the first summand in (E.12), we note that the steps following (E.5) stay valid if we substitute $|\psi_m\rangle \to \Xi|\psi_m\rangle$ or $\langle\psi_n| \to \langle\psi_n|\Xi$, and therefore,

$$\langle\psi_n|\frac{1}{2}\{[P_{\mathbf{r}}, H], \Xi\}|\psi_m\rangle = -\nabla \cdot \mathbf{J}^{(\Xi)}_{nm} \tag{E.13}$$

with

$$\mathbf{J}^{(\Xi)}_{nm} = \frac{1}{4}\langle\psi_n|\{\{P_{\mathbf{r}}, \boldsymbol{\mathcal{V}}\}, \Xi\}|\psi_m\rangle = \frac{1}{4}\langle\psi_n|\{\{\Xi, \boldsymbol{\mathcal{V}}\}, P_{\mathbf{r}}\}|\psi_m\rangle. \tag{E.14}$$

The matrix element with $n = m$ again gives the Ξ-current $\mathbf{J}^{(\Xi)}$, which is real. To summarize, the derivation holds as long as on operator Ξ fulfils the conditions $[\Xi, \hat{\mathbf{r}}] = 0$ and $[\Xi, \hat{\mathbf{k}}] = 0$, since otherwise, the derivative would also act on Ξ.

Finally, we note that the average Ξ-torque, when we evaluate it with an energy eigenstate and integrate over both coordinates of a lead, vanishes. For this, we simply need to use $\int dx \int dy\, P_{\mathbf{r}} = 1$. Then, $\int dx \int dy\, \mathcal{T}_\Xi = \langle\psi|\frac{1}{i}[\Xi, H]|\psi\rangle = 0$ since $H|\psi\rangle = E|\psi\rangle$. Applying this to $\Xi = h_x$, we see that the average helicity current is a conserved quantity. This result may seem trivial - after all, we expect that any non-conserved local polarization that oscillates because of quantum mechanical interference, will cancel on average. Then, the average value (zero) is of course conserved but not very interesting. However, this kind of cancellation is not expected for currents, since oscillations cannot undo the injection of e.g. charge or spin into a lead. The average helicity current, which has been calculated numerically in Chapter 10 is clearly non-zero, electrically detectable and it is also related to a spin-current.

Appendix F

Wave matching for lattice model[1]

We consider the infinite N-SO or SO-SO interface of Chapter 10.3. By employing an approximation of the analytical model by a lattice, we prevent issues with symmetrization [Win+93] and avoid problems with unphysical spurious solutions [Sch+85]. We use the Green's function formalism as developed e.g. in [Wim08] for a tight-binding model with only nearest neighbour hoppings. A shortened derivation of the formalism has also been given in Chapter 5. Since the formalism is the same as is used also in the finite geometry with attached leads, we may call the left and right sides of the interface two "leads". Due to k_y-conservation at the interface, we only need to solve a 1D chain problem with k_y as parameter, while $\hat{k}_x$ will be discretized on a lattice with lattice constant a. Due to translational invariance by shift of a, solutions $\psi(x)$ will be plane waves with k_x in the first Brillouin zone $[-\pi/a, \pi/a]$. However, the 4-band model Hamiltonian (10.1) contains up to 3rd powers of $\hat{k}_x$, which when discretized, lead to next-nearest neighbour hopping elements. Since the formalism is formulated in nearest-neighbour coupling matrices only, we need to use an enlarged unit cell containing 2 lattice sites (we identify them as sublattice A,B). So for the moment, we only make use of translational invariance by $2a$. The ansatz $\psi(x) = e^{ik_x x}\chi(k_x)$ with a spinor $\chi(k_x)$ of 8 components, leads to the effective Schrödinger equation

$$H(k_x)\chi(k_x) = (H_0 + H_1 e^{2ik_x a} + H_{-1}e^{-2ik_x a})\chi(k_x) = E\chi(k_x) \tag{F.1}$$

with $H_{-1} = H_1^\dagger$. H_0, H_1 and H_{-1} are 8×8 matrices describing the Hamiltonian of an isolated enlarged unit cell and the couplings to the right/left cells.

The band structure of $H(k_x)$ is formally obtained from the band structure of the primitive (i.e. single site) unit cell lattice problem by reducing the Brillouin zone to $[-\pi/(2a), \pi/(2a)]$ in the manner of shifting k_x-values by π/a if necessary. So we have twice the number of bands in order to compensate for just half of the original Brillouin zone. We call the bands that are obtained by shifting (they originally have $|\mathrm{Re}(k_x)| > \pi/(2a)$) anti-bonding, and the other bonding. We denote the components $[\chi]_{s,\alpha}$ of χ with an index $s = 0, 1 = A, B$ for the sublattice and $\alpha = 1, ..., 4$ for the band basis in which $\hat{H}$ is written. For the bonding states we have $[\chi]_{A,\alpha}(k_x) = e^{ik_x a}[\chi]_{B,\alpha}(k_x)$ and for the anti-bonding states, $[\chi]_{A,\alpha}(k_x) = -e^{ik_x a}[\chi]_{B,\alpha}(k_x)$. We can use these relations to find the value of k_x in the primitive unit cell model.

Since we have to resort to hoppings by $2a$ anyway, and we finally use the lattice model as an approximation of the analytical **k**-diagonal model, we also use hoppings up to $2a$ to find the discretization of $\hat{k}_x$ and $\hat{k}_x^2$. The discretization of $\hat{k}_x^2$ is found by fitting the parameters c_j for

[1] This appendix has also been published in [Rot+14].

hopping by ja in the most general symmetric dispersion $E(k_x) = c_0 + c_1\cos(k_x) + c_2\cos(2k_x)$. We find $c_0 = 5/2$, $c_1 = -8/3$ and $c_2 = 1/6$. Likewise, the representation of $\hat{k}_x$ is found by fitting $E(k_x) = d_1\sin(k_x) + d_2\sin(2k_x)$, and we find $d_1 = 4/3$ and $d_2 = -1/6$. This does not make the calculations more difficult, but gives a much better approximation to the continuum model.

We numerically solve (F.1) for a fixed energy E_F and find the modes $\chi_n = \chi(k_n)$. The modes are classified in propagating with $|\lambda| = 1$ and evanescent modes with $|\lambda| \neq 1$, and $\lambda = e^{2ik_x a}$. Further they are classified in right-going which is right-decaying ($|\lambda| < 1$) or right-moving ($|\lambda = 1$ and velocity $v > 0$), and left-going. The velocity of a normalized propagating mode $\chi(k_x)$ is obtained by $v = \chi^\dagger(k_x)\frac{\partial H(k_x)}{\partial k_x}\chi(k_x)$. This relation still holds even if the expression $\psi^\dagger(\mathbf{r})\frac{\partial H(k_x)}{\partial k_x}\psi(\mathbf{r})$ cannot be longer interpreted as local current density. This is the case for Hamiltonians containing powers of $\hat{k}_x$ higher than two [Li+07]. We calculate the helicity of a mode by putting the value of k_x into the analytical expression of $h(\mathbf{k})$, Eq. 10.3, and by evaluating the expectation value just for sublattice A ($[\chi]_A$ is a 4-component vector):

$$\langle\hat{h}\rangle_{\chi(\mathbf{k})} = \frac{[\chi]_A^\dagger h(\mathbf{k})[\chi]_A}{[\chi]_A^\dagger[\chi]_A} \tag{F.2}$$

Here it is essential to put in the k_x-value obtained for the primitive lattice, by identifying the bonding/anti-bonding character of χ. Because the rotational symmetry is broken, we do not expect to have perfect values ±1 for the helicity, but it turns out that the approximation to the analytical model works quite well - the helicity expectation value deviates from ±1 by less than 10^{-6}.

Next, we use the eigenmodes to calculate the self-energies Σ_R, Σ_L of the right and left lead. Of course, we have to calculate the modes separately for left and right lead if the parameters of the Hamiltonian depend on the region. But here we just use modes of the right lead to present formulas for both leads. With notation of [Wim08], "$>$" stands for outgoing and "$<$" for ingoing states, so at the right lead, "$>$" stands for left-going and "$<$" is right-going. The 16 modes are sorted into two matrices, $U_> = (\chi_{1,>}, ..., \chi_{8,>})$ and $U_< = (\chi_{1,<}, ..., \chi_{8,<})$. The corresponding eigenvalues λ_n define matrices $\Lambda_> = \mathrm{diag}(\lambda_{1,>}, ..., \lambda_{8,>})$ and $\Lambda_< = \mathrm{diag}(\lambda_{1,<}, ..., \lambda_{8,<})$. Then we have (compare Eq. (5.90) or [Wim08])

$$\Sigma_R = H_1 U_< \Lambda_< U_<^{-1}, \quad \Sigma_L = H_{-1} U_> \Lambda_>^{-1} U_>^{-1}.$$

We can also obtain Σ_L from Σ_R by a rotation about π. For that, we have to combine rotations acting on band and sublattice space and $k_y \to -k_y$. The corresponding Γ matrices are $\Gamma_R = i(\Sigma_R - \Sigma_R^\dagger)$ and $\Gamma_L = i(\Sigma_L - \Sigma_L^\dagger)$. It is very important to use right-decaying states for the right self energy and left-decaying states for the left self energy (evanescent states constitute the Hermitian part of $\Sigma_{L,R}$). The Γ matrices, on the other hand, project only on the propagating states, and e.g. Γ_R may be rewritten with right-propagating states or left-propagating states.

Finally we need the retarded Green's function $G^R = (E_F - H_0 - \Sigma_R - \Sigma_L)^{-1}$ of a single unit cell with the leads attached, modelled by the self-energies. Here we have the possibility to put in different Rashba SO values for H_0 in order to control the smoothness of the interface, e.g. we may use an average of the left and right lead's SO parameters for a smoothed interface. It turns out that cross-helicity transmissions (see Chapter 10.3) decrease upon making the interface smoother.

The full scattering matrix can be obtained with the generalized Fisher-Lee relations (5.74) [Wim08] for transmission and reflection coefficients. For the left-to-right transmission amplitudes, right-going modes in the left lead are matched with right-going modes of the right lead.

For the left-to-left reflection amplitudes, we match right-going modes with left-going modes in the left lead,

$$t_{L;m,n} = \frac{i}{\sqrt{|v_{R;m,<} v_{L;n,>}|}} \chi^\dagger_{R;m,<} \, \Gamma_R \, G^R \, \Gamma_L \, \chi_{L;n,>} \tag{F.3}$$

$$r_{L;m,n} = \frac{1}{\sqrt{|v_{L;m,<} v_{L;n,>}|}} \chi^\dagger_{L;m,<} \left(i\Gamma_L \, G^R \, \Gamma_L - \Gamma_L \right) \chi_{L;n,>}. \tag{F.4}$$

Bibliography

[Abr88] A. A. Abrikosov. *Fundamentals of the Theory of Metals*. North Holland, 1988.

[Alt+97] Alexander Altland and Martin R. Zirnbauer. "Nonstandard symmetry classes in mesoscopic normal-superconducting hybrid structures". In: *Phys. Rev. B* 55 (Jan. 1997), pp. 1142–1161.

[Alv+92] Santos F. Alvarado and Philippe Renaud. "Observation of spin-polarized-electron tunneling from a ferromagnet into GaAs". In: *Phys. Rev. Lett.* 68 (Mar. 1992), pp. 1387–1390.

[Ama+14] A. Amaricci, J. C. Budich, M. Capone, B. Trauzettel, and G. Sangiovanni. "First order character and observable signatures of topological quantum phase transitions". In: *ArXiv e-prints* (Nov. 2014). arXiv:1411.7390 [cond-mat.str-el].

[Ame+01] P. R. Amestoy, I. S. Duff, J. Koster, and J.-Y. L'Excellent. "A Fully Asynchronous Multifrontal Solver Using Distributed Dynamic Scheduling". In: *SIAM Journal on Matrix Analysis and Applications* 23.1 (2001), pp. 15–41.

[And+99] E. Anderson et al. *LAPACK Users' Guide*. Third edition. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1999. ISBN: 0-89871-447-8.

[And72] P. W. Anderson. "More Is Different". In: *Science* 177.4047 (1972), pp. 393–396.

[And13] Yoichi Ando. "Topological Insulator Materials". In: *Journal of the Physical Society of Japan* 82.10 (2013), p. 102001.

[And+87] Lucio Claudio Andreani, Alfredo Pasquarello, and Franco Bassani. "Hole subbands in strained GaAs-$Ga_{1-x}Al_xAs$ quantum wells: Exact solution of the effective-mass equation". In: *Phys. Rev. B* 36 (1987), pp. 5887–5894.

[Aro+98] Daniel P. Arovas and Yuli Lyanda-Geller. "Non-Abelian geometric phases and conductance of spin-$\frac{3}{2}$ holes". In: *Phys. Rev. B* 57 (May 1998), pp. 12302–12305.

[Asm+13] Mahmoud M. Asmar and Sergio E. Ulloa. "Rashba spin-orbit interaction and birefringent electron optics in graphene". In: *Phys. Rev. B* 87 (Feb. 2013), p. 075420.

[Avr+88] J. E. Avron, L. Sadun, J. Segert, and B. Simon. "Topological Invariants in Fermi Systems with Time-Reversal Invariance". In: *Phys. Rev. Lett.* 61 (1988), pp. 1329–1332.

[Avr+89] J. E. Avron, L. Sadun, J. Segert, and B. Simon. "Chern numbers, quaternions, and Berry's phases in Fermi systems". In: *Comm. Math. Phys.* 124.4 (1989), pp. 595–627. ISSN: 0010-3616.

[Aws+07] D. D. Awschalom and M. E. Flatte. "Challenges for semiconductor spintronics". In: *Nature Phys.* 3 (2007), p. 153.

[Bal11] A. A. Balandin. "Thermal properties of graphene and nanostructured carbon materials". In: *Nature Materials* 10.569 (2011).

[Bar+89] H. U. Baranger and A. D. Stone. "Electrical linear-response theory in an arbitrary magnetic field: A new Fermi-surface formation". In: *Phys. Rev. B* 40.12 (Oct. 1989), pp. 8169–8193.

[Bar08] J. H. Bardarson. "Effects of Spin-Orbit Coupling on Quantum Transport". PhD thesis. Leiden University, 2008.

[Bas88] G. Bastard. *Wave mechanics applied to semiconductor heterostructures*. Monographies de physique. Les Éditions de Physique, 1988. ISBN: 9780470217085.

[Bau11] G. E. W. Bauer. "Spin Caloritronics". In: *ArXiv e-prints* (July 2011). arXiv:1107.4395 [cond-mat.mes-hall].

[Bee+91] C. W. J. Beenakker and H. van Houten. "Quantum Transport in Semiconductor Nanostructures". In: Solid State Physics 44 (1991). Ed. by Henry Ehrenreich and David Turnbull, pp. 1 –228. ISSN: 0081-1947.

[Bee+92] C. W. Beenakker and A. A. M. Staring. "Theory of the thermopower of a quantum dot". In: *Physical Review B* 46.9667 (1992).

[Beh+07] K. Behnia, Marie-Aude Méasson, and Y. Kopelevich. "Oscillating Nernst-Ettingshausen effect in Bismuth accross the quantum limit". In: *Physical Review Letters* 98.166602 (2007).

[Bel08] L. E. Bell. "Cooling, heating, generating power, and recoring waste heat with thermoelectric systems". In: *Science* 321.1457 (2008).

[Bel+04] R. Bel et al. "Giant Nernst Effect in $CeCoIn_5$". In: *Physical Review Letters* 92.217002 (2004).

[Bér+06] Alain Bérard and Hervé Mohrbach. "Spin Hall effect and Berry phase of spinning particles". In: *Physics Letters A* 352.3 (2006), pp. 190 –195. ISSN: 0375-9601.

[Ber+12] Dario Bercioux, Daniel F. Urban, Francesco Romeo, and Roberta Citro. "Rashba spin-orbit-interaction-based quantum pump in graphene". In: *Applied Physics Letters* 101.12, 122405 (2012), p. 122405.

[Ber+10] D. Bercioux and A. De Martino. "Spin-resolved scattering through spin-orbit nanostructures in graphene". In: *Phys. Rev. B* 81 (Apr. 2010), p. 165410.

[Ber+06a] B. A. Bernevig, T. L. Hughes, and S. C. Zhang. "Quantum Spin Hall Effect and Topological Phase Transition in HgTe Quantum Wells". In: *Science* 314 (2006), p. 1757.

[Ber+06b] B. Andrei Bernevig and Shou-Cheng Zhang. "Quantum Spin Hall Effect". In: *Phys. Rev. Lett.* 96 (Mar. 2006), p. 106802.

[Ber+05] B. A. Bernevig and Shou-Cheng Zhang. "Holonomic quantum computing based on the Stark effect". In: *Phys. Rev. B* 71 (Jan. 2005), p. 035303.

[Ber84] M. V. Berry. "Quantal Phase Factors Accompanying Adiabatic Changes". In: *Proc. R. Soc. Lond.* 392 (1984), pp. 45–57.

[Bir+63] G. L. Bir, E. I. Butikov, and G. E. Pikus. "Spin and combined resonance on acceptor centres in Ge and Si type crystals-II: The effect of the electrical field and relaxation time". In: *Journal of Physics and Chemistry of Solids* 24.12 (1963), pp. 1475 –1486.

[Bi+13] Xintao Bi, Peiru He, E. M. Hankiewicz, R. Winkler, Giovanni Vignale, and Dimitrie Culcer. "Anomalous spin precession and spin Hall effect in semiconductor quantum wells". In: *Phys. Rev. B* 88 (July 2013), p. 035316.

[Bor+28] M. Born and V. Fock. "Beweis des Adiabatensatzes". German. In: *Zeitschrift für Physik* 51.3-4 (1928), pp. 165–180. ISSN: 0044-3328.

[Bro+10a] Valentina Brosco, Markus Jerger, Pablo San-José, Gergely Zarand, Alexander Shnirman, and Gerd Schön. "Prediction of resonant all-electric spin pumping with spin-orbit coupling". In: *Phys. Rev. B* 82 (July 2010), p. 041309.

[Bro+10b] V. Brosco, M. Jerger, P. San-José, G. Zarand, A. Shnirman, and G. Schön. "Prediction of resonant all-electric spin pumping with spin-orbit coupling". In: *Phys. Rev. B* 82.4 (2010), p. 041309.

[Brü+12] Christoph Brüne et al. "Spin polarization of the quantum spin Hall edge states". In: *Nat. Phys.* 8 (2012), p. 486.

[Brü+10] C. Brüne et al. "Evidence for the ballistic intrinsic Spin-Hall Effect in HgTe nanostructures". In: *Nature Physics* 6 (2010), p. 448.

[Bru+09] Daniel Brunner et al. "A Coherent Single-Hole Spin in a Semiconductor". In: *Science* 325.5936 (2009), pp. 70–72.

[Bud12] Jan Carl Budich. "Fingerprints of Geometry and Topology on Low Dimensional Mesoscopic Systems". PhD thesis. Universität Würzburg, 2012.

[Bud+12a] J. C. Budich, D. G. Rothe, E. M. Hankiewicz, and B. Trauzettel. "All-electric qubit control in heavy hole quantum dots via non-Abelian geometric phases". In: *Phys. Rev. B* 85 (May 2012), p. 205425.

[Bud+12b] J. C. Budich and B. Trauzettel. "Local topological phase transitions in periodic condensed matter systems". In: *The European Physical Journal B* 85.3, 94 (2012).

[Bul+08] Denis V. Bulaev, Björn Trauzettel, and Daniel Loss. "Spin-orbit interaction and anomalous spin relaxation in carbon nanotube quantum dots". In: *Phys. Rev. B* 77 (June 2008), p. 235301.

[Bul+05] D. V. Bulaev and D. Loss. "Spin Relaxation and Decoherence of Holes in Quantum Dots". In: *Physical Review Letters* 95.7, 076805 (Aug. 2005), p. 076805.

[Bur88a] M. G. Burt. "A new effective-mass equation for microstructures". In: *Semiconductor Science and Technology* 3.12 (1988), p. 1224.

[Bur88b] M. G. Burt. "An exact formulation of the envelope function method for the determination of electronic states in semiconductor microstructures". In: *Semicond. Sci. Technol.* 3 (1988), p. 739.

[Bur99] M. G. Burt. "Fundamentals of envelope function theory for electronic states and photonic modes in nanostructures". In: *Journal of Physics: Condensed Matter* 11.9 (1999), p. 53.

[Büt+85] M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas. "Generalized many-channel conductance formula with application to small rings". In: *Phys. Rev. B* 31 (May 1985), pp. 6207–6215.

[Byc+84] Y. A. Bychkov and E. I. Rashba. "Oscillatory effects and the magnetic susceptibility of carriers in inversion layers". In: *J. Phys. C: Solid State Phys.* 17 (1984), p. 6039.

[Cha+10] J. Chang et al. "Nernst and Seebeck Coefficients of the Cuprate Superconductor $YBa_2Cu_3O_{6.67}$: A Study of Fermi Surface Reconstruction". In: *Physical Review Letters* 104.057005 (2010).

[Chu91] Shun Lien Chuang. "Efficient band-structure calculations of strained quantum wells". In: *Phys. Rev. B* 43 (1991), pp. 9649–9661.

[Chu95] Shun Lien Chuang. *Physics of Optoelectronic Devices*. Wiley-Interscience, 1995.

[Chu+07] J. Chu and A. Sher. *Physics and Properties of Narrow Gap Semiconductors*. Microdevices. Springer, 2007. ISBN: 9780387748016.

[Chu+10] Chih-Piao Chuu, Ming-Che Chang, and Qian Niu. "Semiclassical dynamics and transport of the Dirac spin". In: *Solid State Communications* 150.11–12 (2010), pp. 533–537. ISSN: 0038-1098.

[Cul+05] Dimitrie Culcer, Yugui Yao, and Qian Niu. "Coherent wave-packet evolution in coupled bands". In: *Phys. Rev. B* 72 (Aug. 2005), p. 085110.

[Cut+69] Melvin Cutler and N. F. Mott. "Observation of Anderson Localization in an Electron Gas". In: *Phys. Rev.* 181 (May 1969), pp. 1336–1340.

[Czy07] Gerd Czycholl. *Theoretische Festkörperphysik: Von den klassischen Modellen zu modernen Forschungsthemen*. Springer-Lehrbuch. Springer, 2007. ISBN: 9783540747895.

[Dai+08] Xi Dai, Taylor L. Hughes, Xiao-Liang Qi, Zhong Fang, and Shou-Cheng Zhang. "Helical edge and surface states in HgTe quantum wells and bulk insulators". In: *Phys. Rev. B* 77 (Mar. 2008), p. 125319.

[Dat05] S. Datta. *Quantum Transport: Atom to Transistor*. Cambridge University Press, 2005.

[Dat07] S. Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, 2007.

[deGre+11] K. de Greve et al. "Ultrafast coherent control and suppressed nuclear feedback of a single quantum dot hole qubit". In: *Nature Physics* 7 (2011), pp. 872–878.

[Del+12] Pierre Delplace, Jian Li, and Markus Büttiker. "Magnetic-Field-Induced Localization in 2D Topological Insulators". In: *Phys. Rev. Lett.* 109 (Dec. 2012), p. 246803.

[Dem+99] Eugene Demler and Shou-Cheng Zhang. "Non-Abelian Holonomy of BCS and SDW Quasiparticles". In: *Annals of Physics* 271.1 (1999), pp. 83 –119. ISSN: 0003-4916.

[Dir31] P.A.M. Dirac. "Quantised Singularities in the Electromagnetic Field". In: *Proc. Roy. Soc. A* 133 (1931), p. 60.

[Dre55] G. Dresselhaus. "Spin-Orbit Coupling Effects in Zinc Blende Structures". In: *Phys. Rev.* 100 (Oct. 1955), pp. 580–586.

[Dre+08] M. S. Dresselhaus, G. Dresselhaus, and A. Jorio. *Group Theory: Application to the Physics of Condensed Matter*. Springer, 2008.

[Dub+11] Y. Dubi and M. Di Ventra. "Colloquium: Heat flow and thermoelectricity in atomic and molecular junctions". In: *Review of Modern Physics* 83.131 (2011).

[Dya+71] M. I. Dyakonov and V. I. Perel. "Current-induced spin orientation of electrons in semiconductors". In: *Phys. Lett.* A35 (1971), p. 459.

[Dyr+12] A. Dyrdał and J. Barnaś. "Intrinsic contribution to spin Hall and spin Nernst effects in a bilayer graphene". In: *Journal of Physics: Condensed Matter* 24.27 (2012), p. 275302.

[Ebl+09] B. Eble et al. "Hole-Nuclear Spin Interaction in Quantum Dots". In: *Phys. Rev. Lett.* 102 (Apr. 2009), p. 146601.

[Eco06] E.N. Economou. *Green's Functions in Quantum Physics*. Springer Series in Solid-State Sciences. Springer, 2006. ISBN: 9783540288411.

[Eke+85] U. Ekenberg and M. Altarelli. "Subbands and Landau levels in the two-dimensional hole gas at the GaAs-$Al_xGa_{1-x}As$ interface". In: *Phys. Rev. B* 32 (Sept. 1985), pp. 3712–3722.

[Fey82] Richard P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21.6-7 (1982), pp. 467–488.

[Fis+10] Jan Fischer and Daniel Loss. "Hybridization and Spin Decoherence in Heavy-Hole Quantum Dots". In: *Phys. Rev. Lett.* 105 (2010), p. 266603.

[Fis+81] D. S. Fisher and P. A. Lee. "Relation between conductivity and transmission matrix". In: *Phys. Rev. B* 23.12 (June 1981), pp. 6851–6854.

[Fol+50] Leslie L. Foldy and Siegfried A. Wouthuysen. "On the Dirac Theory of Spin $\frac{1}{2}$ Particles and its Non-Relativistic Limit". In: *Phys. Rev.* 78 (Apr. 1950), pp. 29–36.

[For93] B. A. Foreman. "Effective-mass Hamiltonian and boundary conditions for the valence bands of semiconductor microstructures". In: *Phys. Rev. B* 48.7 (Aug. 1993), pp. 4964–4967.

[For97] Bradley A. Foreman. "Elimination of spurious solutions from eight-band k.p theory". In: *Phys. Rev. B* 56 (Nov. 1997), R12748–R12751.

[Fra+13] M. Franz and L. Molenkamp. *Topological Insulators*. Contemporary Concepts of Condensed Matter Science. Elsevier Science, 2013. ISBN: 9780444633187.

[Fu11] Liang Fu. "Topological Crystalline Insulators". In: *Phys. Rev. Lett.* 106 (Mar. 2011), p. 106802.

[Fu+07a] Liang Fu and C. L. Kane. "Topological insulators with inversion symmetry". In: *Phys. Rev. B* 76 (July 2007), p. 045302.

[Fu+08] Liang Fu and C. L. Kane. "Superconducting Proximity Effect and Majorana Fermions at the Surface of a Topological Insulator". In: *Phys. Rev. Lett.* 100 (Mar. 2008), p. 096407.

[Fu+07b] Liang Fu, C. L. Kane, and E. J. Mele. "Topological Insulators in Three Dimensions". In: *Phys. Rev. Lett.* 98 (Mar. 2007), p. 106803.

[Gar+10] E. S. Garlid, Q. O. Hu, M. K. Chan, C. J. Palmstrøm, and P. A. Crowell. "Electrical Measurement of the Direct Spin Hall Effect in Fe/$In_xGa_{1-x}As$ Heterostructures". In: *Physical Review Letters* 105.156602 (2010).

[Ger+08] Brian D. Gerardot et al. "Optical pumping of a single hole spin in a quantum dot". In: *Nature* 451 (2008), pp. 441–444.

[Gia+06] F. Giazotto, T. T. Heikillä, A. Luukanen, A. M. Savin, and J. P. Pekola. "Opportunities for mesoscopics in thermometry and refrigeration: Physics and application". In: *Review of Modern Physics* 78.217 (2006).

[Gol+06] Vitaly N. Golovach, Massoud Borhani, and Daniel Loss. "Electric-dipole-induced spin resonance in quantum dots". In: *Phys. Rev. B* 74 (2006), p. 165319.

[Gol+10] Vitaly N. Golovach, Massoud Borhani, and Daniel Loss. "Holonomic quantum computation with electron spins in quantum dots". In: *Phys. Rev. A* 81 (Feb. 2010), p. 022315.

[Gov+03] M. Governale, F. Taddei, and Rosario Fazio. "Pumping spin with electrical fields". In: *Phys. Rev. B* 68 (Oct. 2003), p. 155324.

[Gui+11] M. Guigou, P. Recher, J. Cayssol, and B. Trauzettel. "Spin Hall effect at interfaces between HgTe/CdTe quantum wells and metals". In: *Phys. Rev. B* 84 (Sept. 2011), p. 094534.

[Gur11] V. Gurarie. "Single-particle Green's functions and interacting topological insulators". In: *Phys. Rev. B* 83 (Feb. 2011), p. 085426.

[Hal82] B. I. Halperin. "Quantized Hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential". In: *Phys. Rev. B* 25 (Feb. 1982), pp. 2185–2190.

[Han+08a] N. Hanasaki et al. "Anomalous Nernst Effects in Pyrochlore Molybdates with Spin Chirality". In: *Physical Review Letters* 100.106601 (2008).

[Han+05] E. M. Hankiewicz, J. Li, T. Jungwirth, Q. Niu, S.-Q. Shen, and J. Sinova. "Charge Hall effect driven by spin-dependent chemical potential gradients and Onsager relations in mesoscopic systems". In: *Phys. Rev. B* 72.15 (Oct. 2005), p. 155305.

[Han+04] E. M. Hankiewicz, L. W. Molenkamp, T. Jungwirth, and J. Sinova. "Manifestation of the spin Hall effect through charge-transport in the mesoscopic regime". In: *Phys. Rev. B* 70.24 (Dec. 2004), p. 241301.

[Han+06] E. M. Hankiewicz and G. Vignale. "Coulomb corrections to the extrinsic spin-Hall effect of a two-dimensional electron gas". In: *Phys. Rev. B* 73 (Mar. 2006), p. 115339.

[Han+08b] E. M. Hankiewicz and G. Vignale. "Phase Diagram of the Spin Hall Effect". In: *Physical Review Letters* 100.2, 026602 (2008), p. 026602.

[Han+09a] E. M. Hankiewicz and G. Vignale. "Spin-Hall effect and spin-Coulomb drag in doped semiconductors". In: *Journal of Physics: Condensed Matter* 21.25 (2009), 253202 (16pp).

[Han+09b] E. M. Hankiewicz and G. Vignale. "Spin-Hall effect and spin-Coulomb drag in doped semiconductors". In: *Journal of Physics: Condensed Matter* 21.25 (2009), p. 253202.

[Her+63] Frank Herman, Charles D. Kuglin, Kermit F. Cuff, and Richard L. Kortum. "Relativistic Corrections to the Band Structure of Tetrahedrally Bonded Semiconductors". In: *Phys. Rev. Lett.* 11 (Dec. 1963), pp. 541–545.

[Hir99] J. E. Hirsch. "Spin Hall Effect". In: *Phys. Rev. Lett.* 83.9 (Aug. 1999), pp. 1834–1837.

[Kan+05a] C. L. Kane and E. J. Mele. "Quantum Spin Hall Effect in Graphene". In: *Phys. Rev. Lett.* 95 (Nov. 2005), p. 226801.

[Kan+05b] C. L. Kane and E. J. Mele. "Z_2 Topological Order and the Quantum Spin Hall Effect". In: *Phys. Rev. Lett.* 95 (Sept. 2005), p. 146802.

[Kan57] E. O. Kane. "Band structure of indium antimonide". In: *Journal of Physics and Chemistry of Solids* 1.4 (1957), pp. 249–261. ISSN: 0022-3697.

[Kat50] Tosio Kato. "On the Adiabatic Theorem of Quantum Mechanics". In: *Journal of the Physical Society of Japan* 5.6 (1950), pp. 435–439.

[Kat+04] Y. Kato, R. C. Myers, A. C. Gossard, and D. D. Awschalom. "Observation of the Spin Hall Effect in Semiconductors". In: *Science* 306 (2004), p. 1910.

[Kel65] L. V. Keldysh. "Diagram technique for nonequilibrium processes". In: *Sov. Phys. JETP* 20.4 (1965), pp. 1018–1026.

[Kho+04] M. Khodas, A. Shekhter, and A. M. Finkel'stein. "Spin Polarization of Electrons by Nonmagnetic Heterostructures: The Basics of Spin Optics". In: *Phys. Rev. Lett.* 92 (Feb. 2004), p. 086602.

[Kis+98] M. V. Kisin, B. L. Gelmont, and S. Luryi. "Boundary-condition problem in the Kane model". In: *Phys. Rev. B* 58 (Aug. 1998), pp. 4605–4616.

[Kli+11] Jelena Klinovaja, Manuel J. Schmidt, Bernd Braunecker, and Daniel Loss. "Carbon nanotubes in electric and magnetic fields". In: *Phys. Rev. B* 84 (Aug. 2011), p. 085452.

[Kli+80] K. v. Klitzing, G. Dorda, and M. Pepper. "New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance". In: *Phys. Rev. Lett.* 45.6 (Aug. 1980), pp. 494–497.

[Kne+11] Ivan Knez, Rui-Rui Du, and Gerard Sullivan. "Evidence for Helical Edge Modes in Inverted InAs/GaSb Quantum Wells". In: *Phys. Rev. Lett.* 107 (Sept. 2011), p. 136603.

[Koh85] Mahito Kohmoto. "Topological invariant and the quantization of the Hall conductance". In: *Annals of Physics* 160.2 (1985), pp. 343 –354. ISSN: 0003-4916.

[Kön+06] M. König et al. "Direct Observation of the Aharonov-Casher Phase". In: *Phys. Rev. Lett.* 96 (Feb. 2006), p. 076804.

[Kön+07] M. König et al. "Quantum Spin Hall Insulator State in HgTe Quantum Wells". In: *Science* 318.5851 (2007), pp. 766–770.

[Kön+08] M. König et al. "The Quantum Spin Hall Effect: Theory and Experiment". In: *Journal of the Physical Society of Japan* 77.3 (2008), p. 031007.

[Kru+11] Viktor Krueckl and Klaus Richter. "Switching Spin and Charge between Edge States in Topological Insulator Constrictions". In: *Phys. Rev. Lett.* 107 (8 Aug. 2011), p. 086803.

[Kub57] Ryogo Kubo. "Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems". In: *Journal of the Physical Society of Japan* 12.6 (1957), pp. 570–586.

[Lan37] L. Landau. "On the theory of phase transitions (translated title)". In: *Zh. Eksp. Teor. Fiz.* 7 (1937), pp. 19–32.

[Lan+72] David C. Langreth and John W. Wilkins. "Theory of Spin Resonance in Dilute Magnetic Alloys". In: *Phys. Rev. B* 6 (Nov. 1972), pp. 3189–3227.

[Lau81] R. Laughlin. "Quantized Hall conductivity in two dimensions". In: *Phys. Rev. B* 23 (May 1981), pp. 5632–5633.

[Lin+09] Jacob Linder, Takehito Yokoyama, and Asle Sudbø. "Anomalous finite size effects on surface states in the topological insulator Bi_2Se_3". In: *Phys. Rev. B* 80 (Nov. 2009), p. 205401.

[Li+94] Tsung L. Li and Kelin J. Kuhn. "Effects of spin-orbit interaction on the envelope-function equations for semiconductor heterostructures". In: *Phys. Rev. B* 50 (Sept. 1994), pp. 8589–8601.

[Liu+08] Chaoxing Liu, Taylor L. Hughes, Xiao-Liang Qi, Kang Wang, and Shou-Cheng Zhang. "Quantum Spin Hall Effect in Inverted Type-II Semiconductors". In: *Phys. Rev. Lett.* 100 (June 2008), p. 236601.

[Liu+10a] Chao-Xing Liu, Xiao-Liang Qi, HaiJun Zhang, Xi Dai, Zhong Fang, and Shou-Cheng Zhang. "Model Hamiltonian for topological insulators". In: *Phys. Rev. B* 82 (July 2010), p. 045122.

[Liu+10b] Chao-Xing Liu et al. "Oscillatory crossover from two-dimensional to three-dimensional topological insulators". In: *Phys. Rev. B* 81 (Jan. 2010), p. 041307.

[Liu+10c] Xuele Liu and X.C. Xie. "Spin Nernst effect in the absence of a magnetic field". In: *Solid State Communications* 150.11–12 (2010), pp. 471–474. ISSN: 0038-1098.

[Li+07] Yi Li and Ruibao Tao. "Current in a spin-orbit-coupling system". In: *Phys. Rev. B* 75 (Feb. 2007), p. 075319.

[Los+98] Daniel Loss and David P. DiVincenzo. "Quantum computation with quantum dots". In: *Phys. Rev. A* 57 (Jan. 1998), pp. 120–126.

[Low51] Per-Olov Löwdin. "A Note on the Quantum-Mechanical Perturbation Theory". In: *J. Chem. Phys.* 19.11 (1951), pp. 1396–1401.

[Lu+10] Hai-Zhou Lu, Wen-Yu Shan, Wang Yao, Qian Niu, and Shun-Qing Shen. "Massive Dirac fermions and spin physics in an ultrathin film of topological insulator". In: *Phys. Rev. B* 81 (Mar. 2010), p. 115407.

[Lun+05] Anders Mathias Lunde and Karsten Flensberg. "On the Mott formula for the thermopower of non-interacting electrons in quantum point contacts". In: *Journal of Physics: Condensed Matter* 17.25 (2005), p. 3879.

[Lut56] J. M. Luttinger. "Quantum Theory of Cyclotron Resonance in Semiconductors: General Theory". In: *Phys. Rev.* 102.4 (May 1956), pp. 1030–1041.

[Lut+55] J. M. Luttinger and W. Kohn. "Motion of Electrons and Holes in Perturbed Periodic Fields". In: *Phys. Rev.* 97.4 (Feb. 1955), pp. 869–883.

[Mac62] D. K. C. MacDonald. *Thermoelectricity: An Introduction to the Principles*. Wiley, New York, 1962.

[Mac+10] Joseph Maciejko, Xiao-Liang Qi, and Shou-Cheng Zhang. "Magnetoconductance of the quantum spin Hall state". In: *Phys. Rev. B* 82 (Oct. 2010), p. 155310.

[Mei+92] Y. Meir and N. S. Wingreen. "Landauer formula for the current through an interacting electron region". In: *Physical Review Letter* 68.2512 (1992).

[Miy+07] T. Miyasato et al. "Crossover Behavior of the Anomalous Hall Effect and Anomalous Nernst Effect in Itinerant Ferromagnets". In: *Physical Review Letters* 99.086602 (2007).

[Müh+14a] M. Mühlbauer et al. "One-Dimensional Weak Antilocalization Due to the Berry Phase in HgTe Wires". In: *Phys. Rev. Lett.* 112 (Apr. 2014), p. 146803.

[Müh+14b] M. Mühlbauer et al. "Supplementary Online Material to *One-Dimensional Weak Antilocalization due to the Berry Phase in HgTe Wires*". In: *Phys. Rev. Lett.* 112 (Apr. 2014), p. 146803.

[Mur+04] Shuichi Murakami, Naoto Nagosa, and Shou-Cheng Zhang. "SU(2) non-Abelian holonomy and dissipationless spin current in semiconductors". In: *Phys. Rev. B* 69 (June 2004), p. 235206.

[Mur+03] S. Murakami, N. Nagaosa, and S. C. Zhang. "Dissipationless Quantum Spin Current at Room Temperature". In: *Science* 301.5638 (2003), pp. 1348–1351.

[Nak03] M. Nakahara. *Geometry, Topology and Physics, Second Edition*. Graduate student series in physics. Taylor & Francis, 2003. ISBN: 9780750306065.

[Nay+08] Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, and Sankar Das Sarma. "Non-Abelian anyons and topological quantum computation". In: *Rev. Mod. Phys.* 80 (Sept. 2008), pp. 1083–1159.

[Ner87] W. Nernst. "Ueber die electromotorischen Kraefte, welche durch den Magnetismus in von einem Wärmestrome durchflossenen Metallplatten geweckt werden". In: *Annalen der Physik und Chemie* 31.760 (1887).

[Nie+10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. ISBN: 9781139495486.

[Nik+09] B. K. Nikolic, L. P. Zarbo, and S. Souma. "Spin Currents in Semiconductor Nanostructures: A Nonequilibrium Green-Function Approach". In: *arXiv.org:0907.4122* (2009).

[Nik+05a] Branislav K. Nikolić, Liviu P. Zârbo, and Satofumi Souma. "Mesoscopic spin Hall effect in multiprobe ballistic spin-orbit-coupled semiconductor bridges". In: *Phys. Rev. B* 72 (Aug. 2005), p. 075361.

[Nik+05b] Branislav K. Nikolić, Liviu P. Zârbo, and Sven Welack. "Transverse spin-orbit force in the spin Hall effect in ballistic semiconductor wires". In: *Phys. Rev. B* 72 (Aug. 2005), p. 075335.

[Niu+85] Qian Niu, D. J. Thouless, and Yong-Shi Wu. "Quantized Hall conductance as a topological invariant". In: *Phys. Rev. B* 31 (Mar. 1985), pp. 3372–3377.

[Nov+05a] E. G. Novik et al. "Band structure of semimagnetic $Hg_{1-y}Mn_yTe$ quantum wells". In: *Phys. Rev. B* 72.3 (July 2005), p. 035321.

[Nov+05b] K. S. Novoselov et al. "Two-dimensional gas of massless Dirac fermions in graphene". In: *Nature* 438 (2005), pp. 197–200.

[Nov+07] K. S. Novoselov et al. "Room-Temperature Quantum Hall Effect in Graphene". In: *Science* 315.5817 (2007), p. 1379.

[Now+07] K. C. Nowack, F. H. L. Koppens, Yu. V. Nazarov, and L. M. K. Vandersypen. "Coherent Control of a Single Electron Spin with Electric Fields". In: *Science* 318.5855 (2007), pp. 1430–1433.

[Ohn+99] Y. Ohno, D. K. Young, B. Beschoten, F. Matsukura, H. Ohno, and D. D. Awschalom. "Electrical spin injection in a ferromagnetic semiconductor heterostructure". In: *Nature* 402 (1999), pp. 790–792.

[Pfe00] A. Pfeuffer-Jeschke. "Bandstruktur und Landau-Niveaus quecksilberhaltiger II-VI-Heterostrukturen". PhD thesis. Universität Würzburg, 2000.

[Pik+74] G. E. Pikus and G. L. Bir. *Symmetry and Strain-induced Effects in Semiconductors*. Wiley, New York, 1974.

[Rag+08] S. Raghu, Xiao-Liang Qi, C. Honerkamp, and Shou-Cheng Zhang. "Topological Mott Insulators". In: *Phys. Rev. Lett.* 100 (Apr. 2008), p. 156401.

[Ram07] Jørgen Rammer. *Quantum field theory of non-equilibrium states*. Cambridge University Press, 2007.

[Ras03] E. I. Rashba. "Spin currents in thermodynamic equilibrium: The challenge of discerning transport currents". In: *Phys. Rev. B* 68.24 (Dec. 2003), p. 241315.

[Rho+04] J. A. Rhodes and M. D. Semon. "Relativistic velocity space, Wigner rotation, and Thomas precession". In: *Am. J. Phys.* 72.7 (2004).

[Rot+09] Andreas Roth et al. "Nonlocal Transport in the Quantum Spin Hall State". In: *Science* 325.5938 (2009), pp. 294–297.

[Rot+14] D. G. Rothe and E. M. Hankiewicz. "Tunable polarization in a beam splitter based on two-dimensional topological insulators". In: *Phys. Rev. B* 89 (Jan. 2014), p. 035418.

[Rot+12] D. G. Rothe, E. M. Hankiewicz, B. Trauzettel, and M. Guigou. "Spin-dependent thermoelectric transport in HgTe/CdTe quantum wells". In: *Phys. Rev. B* 86 (Oct. 2012), p. 165434.

[Rot+10] D. G. Rothe, R. W. Reinthaler, C.-X. Liu, L. W. Molenkamp, S.-C. Zhang, and E. M. Hankiewicz. "Fingerprint of different spin-orbit terms for spin transport in HgTe quantum wells". In: *New J. of Physics* 12 (2010), p. 065012.

[Rot09] Dietrich G. Rothe. *Diploma thesis: Spin-Hall Effect in Semiconducting Quantum Wells*. Universität Würzburg, 2009.

[Run+08] I. Rungger and S. Sanvito. "Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition". In: *Physical Review B (Condensed Matter and Materials Physics)* 78.3, 035407 (2008), p. 035407.

[Sak04] J. J. Sakurai. *Modern Quantum Mechanics*. Pearson Education, 2004.

[San+08] Pablo San-Jose, Burkhard Scharfenberger, Gerd Schön, Alexander Shnirman, and Gergely Zarand. "Geometric phases in semiconductor spin qubits: Manipulations and decoherence". In: *Phys. Rev. B* 77 (2008), p. 045305.

[San+99] S. Sanvito, C. J. Lambert, J. H. Jefferson, and A. M. Bratkovsky. "General Green's-function formalism for transport calculations with *spd* Hamiltonians and giant magnetoresistance in Co- and Ni-based magnetic multilayers". In: *Phys. Rev. B* 59.18 (May 1999), pp. 11936–11948.

[Sch+00] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip, and B. J. van Wees. "Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor". In: *Phys. Rev. B* 62 (Aug. 2000), R4790–R4793.

[Sch+08] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig. "Classification of topological insulators and superconductors in three spatial dimensions". In: *Phys. Rev. B* 78 (Nov. 2008), p. 195125.

[Sch+09] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig. "Classification of Topological Insulators and Superconductors". In: *AIP Conference Proceedings* 10 (2009), p. 1134.

[Sch+85] M. F. H. Schuurmans and G. W. 't Hooft. "Simple calculations of confinement states in a quantum well". In: *Phys. Rev. B* 31 (June 1985), pp. 8041–8048.

[See26] T. J. Seebeck. "Ueber die magnetische Polarisation der Metalle und Erze durch Temperaturddifferenz". In: *Ann. Phys.* 82 (1826), p. 253.

[Seg05] D. Segal. "Thermoelectric effect in molecular junctions: A tool for revealing transport mechanisms". In: *Physical Review B* 72.165426 (2005).

[Sha+10] Wen-Yu Shan, Hai-Zhou Lu, and Shun-Qing Shen. "Effective continuous model for surface states and thin films of three-dimensional topological insulators". In: *New Journal of Physics* 12.4 (2010), p. 043048.

[She+05] L. Sheng, D. N. Sheng, and C. S. Ting. "Spin-Hall Effect in Two-Dimensional Electron Systems with Rashba Spin-Orbit Coupling and Disorder". In: *Phys. Rev. Lett.* 94.1 (Jan. 2005), p. 016602.

[Shi+06] Junren Shi, Ping Zhang, Di Xiao, and Qian Niu. "Proper Definition of Spin Current in Spin-Orbit Coupled Systems". In: *Phys. Rev. Lett.* 96 (Feb. 2006), p. 076604.

[Sim83] Barry Simon. "Holonomy, the Quantum Adiabatic Theorem, and Berry's Phase". In: *Physical Review Letters* 51.24 (Dec. 12, 1983), pp. 2167–2170.

[Sin08] N. A. Sinitsyn. "Semiclassical theories of the anomalous Hall effect". In: *Journal of Physics: Condensed Matter* 20.2 (2008), p. 023201.

[Sin+04] J. Sinova, D. Culcer, Q. Niu, N. A. Sinitsyn, T. Jungwirth, and A. H. MacDonald. "Universal Intrinsic Spin Hall Effect". In: *Phys. Rev. Lett.* 92.12 (Mar. 2004), p. 126603.

[Sla+11] A. Slachter, F. L. Bakker, and B. J. van Wees. "Anomalous Nernst and anisotropic magnetoresistive heating in a lateral spin valve". In: *Physical Review B* 84.020412(R) (2011).

[Sny97] John Snygg. *Clifford Algebra : A Computational Tool for Physicists*. Oxford University Press, USA, 1997. ISBN: 9780198025863.

[Sol+89] F. Sols, M. Macucci, U. Ravaioli, and K. Hess. "Theory for a quantum modulated transistor". In: *J. Appl. Phys.* 66.8 (1989), p. 3892.

[Sta82] Richard Stacey. "Eliminating lattice fermion doubling". In: *Phys. Rev. D* 26 (July 1982), pp. 468–472.

[Sun+99] Ganesh Sundaram and Qian Niu. "Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and Berry-phase effects". In: *Phys. Rev. B* 59 (June 1999), pp. 14915–14925.

[Sza+89] Aaron Szafer and A. Douglas Stone. "Theory of Quantum Conduction through a Constriction". In: *Phys. Rev. Lett.* 62 (Jan. 1989), pp. 300–303.

[Tak+10] R. Takahashi and S. Murakami. "Thermoelectric transport in perfectly conducting channels in quantum spin Hall systems". In: *Physical Review B* 81 (2010), 161302(R).

[Tho+82] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. "Quantized Hall Conductance in a Two-Dimensional Periodic Potential". In: *Phys. Rev. Lett.* 49 (Aug. 1982), pp. 405–408.

[Tka+10] G. Tkachov and E. M. Hankiewicz. "Ballistic Quantum Spin Hall State and Enhanced Edge Backscattering in Strong Magnetic Fields". In: *Phys. Rev. Lett.* 104 (Apr. 2010), p. 166803.

[Tre+10] O. A. Tretiakov, Ar. Abanov, S. Murakami, and J. Sinova. "Large thermoelectric figure of merit for three-dimensional topological Anderson insulators via line dislocation engineering". In: *Applied Physics Letters* 97 (2010), p. 073108.

[Tre+11] O. A. Tretiakov, Ar. Abanov, and J. Sinova. "Holey topological thermoelectrics". In: *Applied Physics Letters* 99.113110 (2011).

[Tse+06] Wang-Kong Tse and S. Das Sarma. "Intrinsic spin Hall effect in the presence of extrinsic spin-orbit scattering". In: *Phys. Rev. B* 74 (Dec. 2006), p. 245309.

[Two+08] J. Tworzydło, C. W. Groth, and C. W. J. Beenakker. "Finite difference method for transport properties of massless Dirac fermions". In: *Phys. Rev. B* 78 (Dec. 2008), p. 235438.

[Val+06] S. O. Valenzuela and M. Tinkham. "Direct electronic measurement of the spin Hall effect". In: *Nature* 442 (2006), pp. 176–179.

[Vol09] G. E. Volovik. *The Universe in a Helium Droplet*. International Series of Monographs on Physics. OUP Oxford, 2009. ISBN: 9780199564842.

[Wan+13] Zhong Wang and Binghai Yan. "Topological Hamiltonian as an exact tool for topological invariants". In: *Journal of Physics: Condensed Matter* 25.15 (2013), p. 155601.

[Wat+03] Susan K. Watson, R. M. Potok, C. M. Marcus, and V. Umansky. "Experimental Realization of a Quantum Spin Pump". In: *Phys. Rev. Lett.* 91 (Dec. 2003), p. 258301.

[Web+85] R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz. "Observation of $\frac{h}{e}$ Aharonov-Bohm Oscillations in Normal-Metal Rings". In: *Phys. Rev. Lett.* 54 (June 1985), pp. 2696–2699.

[Wee+88] B. J. van Wees et al. "Quantized conductance of point contacts in a two-dimensional electron gas". In: *Phys. Rev. Lett.* 60.9 (Feb. 1988), pp. 848–850.

[Wil+84] Frank Wilczek and A. Zee. "Appearance of Gauge Structure in Simple Dynamical Systems". In: *Phys. Rev. Lett.* 52 (June 1984), pp. 2111–2114.

[Wim08] M. Wimmer. "Quantum transport in nanostructures: From computational concepts to spintronics in graphene and magnetic tunnel junctions". PhD thesis. Universität Regensburg, 2008.

[Win05] R. Winkler. *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*. Springer, 2005.

[Win+93] R. Winkler and U. Rössler. "General approach to the envelope-function approximation based on a quadrature method". In: *Phys. Rev. B* 48 (Sept. 1993), pp. 8918–8927.

[Wol+01] S. A. Wolf et al. "Spintronics: A Spin-Based Electronics Vision for the Future". In: *Science* 294.5546 (2001), pp. 1488–1495.

[Wu+06] C. J. Wu, B. A. Bernevig, and S.-C. Zhang. "Helical Liquid and the Edge of Quantum Spin Hall Systems". In: *Physical Review Letters* 96 (2006), p. 106401.

[Wun+05] J. Wunderlich, B. Kaestner, J. Sinova, and T. Jungwirth. "Experimental Observation of the Spin-Hall Effect in a Two-Dimensional Spin-Orbit Coupled Semiconductor System". In: *Phys. Rev. Lett.* 94.4 (Feb. 2005), p. 047204.

[Xia+10] Di Xiao, Ming-Che Chang, and Qian Niu. "Berry phase effects on electronic properties". In: *Rev. Mod. Phys.* 82 (July 2010), pp. 1959–2007.

[Yam+11] Ai Yamakage, Ken-Ichiro Imura, Jérôme Cayssol, and Yoshio Kuramoto. "Interfacial charge and spin transport in $\mathbb{Z}_2$ topological insulators". In: *Phys. Rev. B* 83 (Mar. 2011), p. 125401.

[Yok+09] Takehito Yokoyama, Yukio Tanaka, and Naoto Nagaosa. "Giant Spin Rotation in the Junction between a Normal Metal and a Quantum Spin Hall System". In: *Phys. Rev. Lett.* 102 (Apr. 2009), p. 166801.

[Zak64] J. Zak. "Magnetic Translation Group". In: *Phys. Rev.* 134 (June 1964), A1602–A1606.

[Zan+99] Paolo Zanardi and Mario Rasetti. "Holonomic quantum computation". In: *Physics Letters A* 264.2-3 (1999), pp. 94 –99. ISSN: 0375-9601.

[Zee88] A. Zee. "Non-Abelian gauge structure in nuclear quadrupole resonance". In: *Phys. Rev. A* 38 (July 1988), pp. 1–6.

[Zha+09] Haijun Zhang, Chao-Xing Liu, Xiao-Liang Qi, Xi Dai, Zhong Fang, and Shou-Cheng Zhang. "Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 with a single Dirac cone on the surface". In: *Nat. Phys.* 5 (2009), pp. 438–442.

[Zha+10] Yi Zhang et al. "Crossover of the three-dimensional topological insulator Bi_2Se_3 to the two-dimensional limit". In: *Nat. Phys.* 6 (2010), pp. 584–588.

[Zho+08] Bin Zhou, Hai-Zhou Lu, Rui-Lin Chu, Shun-Qing Shen, and Qian Niu. "Finite Size Effects on Helical Edge States in a Quantum Spin-Hall System". In: *Phys. Rev. Lett.* 101 (Dec. 2008), p. 246807.

[Zim60] J. M. Ziman. *Electrons and Phonons*. Oxford University Press, 1960.

Acknowledgements

There are many people who have helped me in different ways during the past years that I spent at the research group for mesoscopic physics at the University of Würzburg, and here I wish to express my gratitude to all of them. Having a vivid scientific environment, where frequent exchange of ideas is natural and without reservation, is a prerequisite for staying motivated over a long period and cannot be valued high enough. Without the constant scientific and personal support of these people, this thesis would not exist in the present form.

First of all, I want to thank my advisor Prof. Dr. Ewelina Hankiewicz. With her optimistic nature, she often encouraged me to stay focussed, and I also benefited from her excellent connections to the scientific community. I thank the head of our department (Institut für theoretische Physik IV), Prof. Dr. Björn Trauzettel, for showing appreciable interest in keeping the working atmosphere as positive as it is. A am grateful that I could benefit from his excellent scientific leadership. I am also very much indebted to former group member Dr. Jan Budich. I like to remember our long and interesting discussions going deeply into many fields of physics, even if we ended at quantum mechanics most of the time. I very much benefited from his profound knowledge of physics, as well as his personal support. This also holds for current group member Thore Poßke. Throughout my stay at our department, I felt very lucky to share an office with my very cheerful colleague Rolf Reinthaler, who also put much effort into holding the group together by organizing social events. I want to thank some former group members for being great collaborators. I thank Prof. Dr. Chao-Xing Liu for sharing with me his expertise on calculation of HgTe/CdTe band structures. His kind patience, surely not a quality of all bright minds, made it a very pleasant experience to work with him. Further, I thank Dr. Marine Guigou for discussions of thermoelectric effects and for her nice feedback on my ideas, which I often do not value as much myself. I thank Dr. Joerg Schelter for many fruitful discussions about numerical issues. Thanks for providing a positive working atmosphere with overlap to recreational activities goes to current and former group members Jan Böttcher, Dr. François Crépin, Lars Elster, Moritz Fuchs, Florian Geissler, Hans Hettmansperger, Stefan Jürgens, Christian Michel, Dr. Stefan Walter, and the rest of the TP4 group. Having such nice colleagues was always of great value when exploring the locations at conferences. I also thank our secretary Nelly Meier for keeping administrative matters running smoothly, and I appreciate very much her kindness, which is not an employment criterion for administration staff at German universities. Further, I thank Prof. Dr. Hartmut Buhmann and Mathias Mühlbauer for their collaboration in keeping up our funding.

I am also very much indebted to all my friends, to which I count many colleagues, but beyond those, let me mention Dr. Steffen Lother, Dr. Stefan Stodolkowitz and Dr. Jürgen Grahl. They significantly stimulated my personal development and made life outside the realm of physics enjoyable, which can be of high value when scientific progress is hard. Last but not least, I thank my mother Dorothea Rothe for her constant empathy and personal support.

List of publications

The following articles have appeared in peer-reviewed journals:

1. Dietrich G. Rothe and Ewelina M. Hankiewicz.
   Tunable polarization in beam-splitter based on 2D topological insulators
   Phys. Rev. B **89**, 035418 (2014)

2. Dietrich G. Rothe, Ewelina M. Hankiewicz, Björn Trauzettel, and Marine Guigou.
   Spin-dependent thermoelectric transport in HgTe/CdTe quantum wells
   Phys. Rev. B **86**, 165434 (2012)

3. Jan C. Budich, Dietrich G. Rothe, Ewelina M. Hankiewicz, and Björn Trauzettel.
   All-electric qubit control in heavy hole quantum dots via non-Abelian geometric phases
   Phys. Rev. B **85**, 205425 (2012)

4. Dietrich G. Rothe, Rolf W. Reinthaler, Chaoxing Liu, Laurens W. Molenkamp, Shoucheng Zhang, and Ewelina M. Hankiewicz.
   Fingerprint of different spin-orbit terms for spin transport in HgTe quantum wells
   New J. Phys. **12**, 065012 (2010)

The following article is a conference proceedings:

5. Patrik Recher, Alena G. Novik, Rolf W. Reinthaler, Dietrich G. Rothe, Ewelina M. Hankiewicz, and Björn Trauzettel.
   Signatures of topology in ballistic bulk transport of HgTe quantum wells
   Proc. SPIE Vol. **7760**, 776018 (2010)