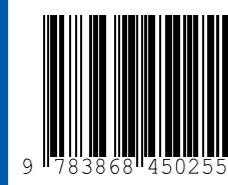


This dissertation deals with advanced computational concepts for quantum transport and spintronics in graphene and magnetic tunnel junctions.

The first part is devoted to developing generic transport algorithms, applicable to any tight-binding Hamiltonian. After a review of Green's function based quantum transport formalisms, the dissertation presents a numerically stable algorithm for computing the surface Green's function of a lead, as well as a generalized Fisher-Lee relation. A matrix reordering algorithm is introduced to extend the applicability of established quantum transport algorithms.

The second part discusses spin-dependent transport in magnetic tunnel junctions and graphene nanoribbons, making use of the numerical techniques developed in the first part as well as analytical models. In particular, this includes an investigation of the tunneling anisotropic magnetoresistance in a strong magnetic field, a thorough discussion of the transport in the graphene edge state, and a study of spin currents in rough graphene nanoribbons.

Universitätsverlag Regensburg



ISBN 978-3-86845-025-5

gefördert von:

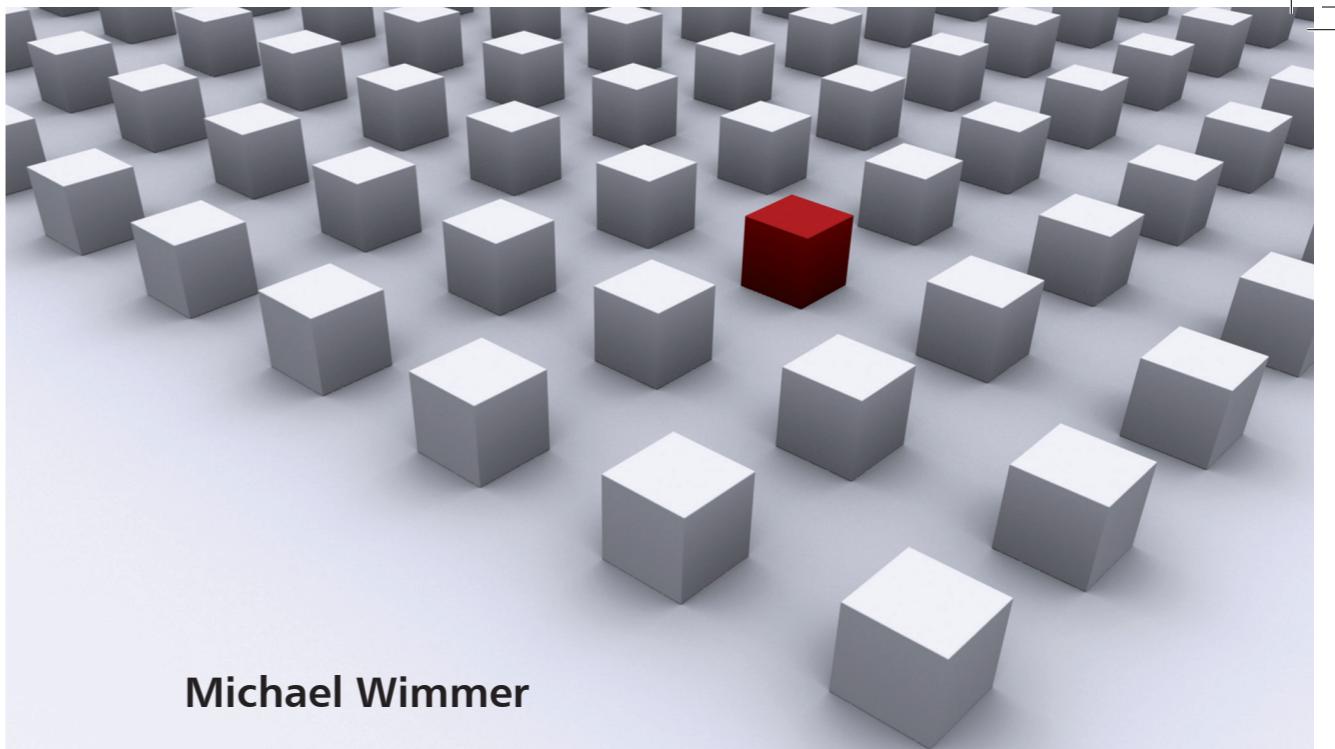


UR

Universität Regensburg

Michael Wimmer

Dissertationsreihe Physik - Band 05



Michael Wimmer

**Quantum transport in nano-
structures: From computational
concepts to spintronics in
graphene and magnetic tunnel
junctions**

Universitätsverlag Regensburg

05

Dissertationsreihe
Physik



Michael Wimmer



Quantum transport in nano-
structures: From computational
concepts to spintronics in
graphene and magnetic tunnel
junctions

Quantum transport in nanostructures: From computational concepts to spintronics in graphene and magnetic tunnel junctions

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der naturwissenschaftlichen Fakultät II - Physik der Universität Regensburg

vorgelegt von

Michael Wimmer

aus Lappersdorf

Oktober 2008

Die Arbeit wurde von Prof. Dr. Klaus Richter angeleitet.

Das Promotionsgesuch wurde am 02.07.2008 eingereicht.

Das Kolloquium fand am 01.10.2008 statt.

Prüfungsausschuss: Vorsitzender: Prof. Dr. Dieter Weiss

1. Gutachter: Prof. Dr. Klaus Richter

2. Gutachter: Prof. Dr. John Schliemann

weiterer Prüfer: Prof. Dr. Gunnar Bali

Ersatzprüfer: Prof. Dr. Vladimir Braun



**Dissertationsreihe der Fakultät für Physik der Universität Regensburg,
Band 05**

Herausgegeben vom Präsidium des Alumnivereins der Physikalischen Fakultät:

Klaus Richter, Andreas Schäfer, Werner Wegscheider

Michael Wimmer

**Quantum transport in nano-
structures: From computational
concepts to spintronics in
graphene and magnetic tunnel
junctions**

Universitätsverlag Regensburg

Bibliografische Informationen der Deutschen Bibliothek.
Die Deutsche Bibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie. Detaillierte bibliografische Daten
sind im Internet über <http://dnb.ddb.de> abrufbar.

1. Auflage 2009

© 2009 Universitätsverlag, Regensburg
Leibnizstraße 13, 93055 Regensburg

Konzeption: Thomas Geiger

Umschlagentwurf: Franz Stadler, Designcooperative Nittenau eG

Layout: Michael Wimmer

Druck: Docupoint, Magdeburg

ISBN: 978-3-86845-025-5

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlags ist es
nicht gestattet, dieses Buch oder Teile daraus auf fototechnischem oder
elektronischem Weg zu vervielfältigen.

Weitere Informationen zum Verlagsprogramm erhalten Sie unter:
www.univerlag-regensburg.de

**Quantum transport in nanostructures:
From computational concepts to spintronics
in graphene and magnetic tunnel junctions**



DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT II - PHYSIK

DER UNIVERSITÄT REGENSBURG

vorgelegt von

Michael Wimmer

aus
Lappersdorf

im Jahr 2008

Promotionsgesuch eingereicht am: 02.07.2008

Die Arbeit wurde angeleitet von: Prof. Dr. Klaus Richter

Prüfungsausschuss: Vorsitzender: Prof. Dr. Dieter Weiss

1. Gutachter: Prof. Dr. Klaus Richter

2. Gutachter: Prof. Dr. John Schliemann

weiterer Prüfer: Prof. Dr. Gunnar Bali

Ersatzprüfer: Prof. Dr. Vladimir Braun

Quantum transport in nanostructures: From computational concepts to spintronics in graphene and magnetic tunnel junctions



Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)
der Naturwissenschaftlichen Fakultät II – Physik
der Universität Regensburg

vorgelegt von
Michael Wimmer
aus Lappersdorf

2008

Promotionsgesuch eingereicht am 2. Juli 2008
Promotionskolloquium am 1. Oktober 2008



Die Arbeit wurde angeleitet von Prof. Dr. Klaus Richter

Prüfungsausschuß:

- | | |
|------------------|---------------------------|
| Vorsitzender: | Prof. Dr. Dieter Weiss |
| 1. Gutachter: | Prof. Dr. Klaus Richter |
| 2. Gutachter: | Prof. Dr. John Schliemann |
| Weiterer Prüfer: | Prof. Dr. Gunnar Bali |
| Ersatzprüfer: | Prof. Dr. Vladimir Braun |

Contents

1	Introduction	1
1.1	Beyond Moore's law	1
1.2	The need for numerics	4
1.3	Outline	5
I	Computational concepts: A generic approach to transport in tight-binding models	9
2	Green's function formalism for transport	13
2.1	Introduction	13
2.2	Basic definitions	14
2.3	Green's functions of non-interacting systems	19
2.4	Non-equilibrium perturbation theory	24
2.4.1	The need for a non-equilibrium theory	24
2.4.2	Contour-ordered Green's function theory	26
2.4.3	Real-time formulation	30
2.5	Transport in tight-binding models	33
2.5.1	Green's functions in tight-binding	33
2.5.2	Transport equations	34
2.5.3	Connection with the scattering formalism	39
2.6	Challenges for a numerical implementation	43
3	Lead Green's functions	45
3.1	Introduction	45
3.2	A general expression for the lead Green's function	48
3.2.1	Green's function of the infinite wire	48
3.2.2	Zeroes of $(E + i\eta - H_0)z - H_1 z^2 - H_{-1}$	50
3.2.3	The role of degenerate modes	53

3.2.4	Solving the contour integral	55
3.3	A more stable algorithm based on the Schur decomposition	57
3.3.1	Failure of the eigendecomposition based algorithm	57
3.3.2	Surface Green's function from the Schur decomposition	61
3.3.3	Numerical Stability	64
3.3.4	Examples	65
3.4	Summary	66
4	Optimal block-tridiagonalization of matrices for quantum transport	69
4.1	Introduction	69
4.2	Definition of the problem	72
4.2.1	Definition of the matrix reordering problem	72
4.2.2	Mapping onto a graph partitioning problem	73
4.3	Optimal Matrix reordering by graph partitioning	75
4.3.1	A local approach—breadth first search	75
4.3.2	A global approach—recursive bisection	77
4.3.3	Computational complexity	83
4.4	Examples: Charge transport in two-dimensional systems	83
4.4.1	Ballistic transport in two-terminal devices	83
4.4.2	Multi-terminal structures	90
4.5	Summary	92
II	Spintronics in graphene and magnetic tunnel junctions	95
5	Magnetic field effects on tunneling anisotropic magnetoresistance	99
5.1	Introduction	99
5.2	Magnetic field dependence of TAMR	101
5.2.1	Model	101
5.2.2	Numerical results	105
5.2.3	A qualitative picture	105
5.3	Summary	111
6	The graphene edge state	113
6.1	Introduction	113
6.2	Fundamentals of graphene	114
6.2.1	Graphene Hamiltonian	114
6.2.2	Symmetries of the graphene Hamiltonian	116
6.3	Characterization of the edge state	118
6.3.1	Edge state at a single zigzag edge	118
6.3.2	Edge state in a zigzag nanoribbon	123
6.3.3	Effects of a staggered potential	126
6.3.4	Effects of next-nearest neighbor hopping	128

6.4	Edge state transport in zigzag nanoribbons	131
6.4.1	Model-dependence of edge state transport—some peculiar examples	131
6.4.2	Where does the current flow?	135
6.4.3	Effects of perturbations on the current flow	139
6.4.4	Scattering from edge defects	145
6.5	Summary	148
7	Edge state based spintronics in graphene	149
7.1	Introduction	149
7.2	Mean-field model for edge-state magnetism	151
7.3	Spin currents in rough graphene nanoribbons	154
7.3.1	Basic ideas	154
7.3.2	Spin conductance of rough graphene nanoribbons	155
7.3.3	Universal spin conductance fluctuations	160
7.3.4	All-electrical detection of edge magnetism	164
7.4	Edge-state induced Spin Hall effect	165
7.4.1	Spin-dependent deflection of the edge state	165
7.4.2	Pure spin currents in graphene micro-bridges	168
7.5	Summary	172
8	Summary and perspectives	173
8.1	Summary	173
8.2	Outlook	175
Appendix		177
A	Observables in tight-binding	177
A.1	Single-particle operators in tight-binding	177
A.2	Observables in terms of Green's functions	179
B	The recursive Green's function method	183
B.1	Basic principles	183
B.2	Summary of algorithms	187
B.3	Computational complexity and implementation	189
C	Generalized Fisher-Lee relation	191
C.1	Landauer-Büttiker formalism	191
C.2	Generalized orthogonality relations	193
C.3	Fisher-Lee relation	196
C.4	Equivalence of scattering and NEGF formalism	199

D Details of the derivation of the lead Green's function	201
D.1 $\lambda(E + i\eta)$ for propagating modes	201
D.2 Derivation of Eq. (3.27)	203
D.3 Summary of the numerical algorithms	205
D.3.1 Eigendecomposition based algorithms	205
D.3.2 Schur decomposition based algorithms	206
E The Fiduccia-Mattheyses algorithm	209
E.1 Graphs and hypergraphs	209
E.2 Fiduccia-Mattheyses bisection	210
F The method of finite differences	211
F.1 Basic ideas	211
F.2 Example	212
F.3 Finite differences form of the Hamiltonian	213
F.4 Bloch's theorem and periodic boundary conditions	215
G Tight-binding model for graphene	217
G.1 Lattice structure	217
G.2 Electronic structure	219
G.2.1 Tight-binding model	219
G.2.2 Band structure in tight-binding approximation	222
G.2.3 Effective Hamiltonian	226
G.3 Graphene nanoribbons	229
G.3.1 Band structure	229
G.3.2 Boundary conditions	230
References	233

*An jeder Sache etwas zu sehen suchen,
was noch niemand gesehen
und woran noch niemand gedacht hat.*

Georg Christoph Lichtenberg (1742–1799)

Chapter 1

Introduction

1.1 Beyond Moore's law

Nowadays, electronic devices are abundant in our daily lives. Electronics has created previously unthought possibilities for communication and information processing in the form of computers and communication devices, but has also found its way into everyday items, such as telephones or cars—for the better or the worse. A few decades ago, a simple table-top calculator used to be a high-end device, but today even a coffee machine can have more computing power. Indeed, the field of electronics has evolved tremendously in the second half of the last century.

This development is commonly visualized in the form of *Moore's law*: In 1965, Gordon Moore predicted that the number of transistors on a chip would double every year [1]. Although Moore was only bold enough to make a prediction for the next ten years, and chips at that time only involved a few tens to a hundred transistors, remarkably this exponential growth has continued until today, with a few billion transistors in high-end computer processors. This exponential growth is best presented as a picture: In Fig. 1.1 we show Moore's law on the example of the number of transistors in Intel processors as a function of time, together with Moore's original prediction. Indeed, we observe a doubling of the number of transistors approximately every 18 months, valid in the last 40 years since Moore's prediction. However, as any exponential growth, this increase must end at some point.

Increasing the performance of electronic devices was always driven by scaling down the system sizes. On the way from dimensions of a few millimeters to a few ten nanometers, conventional semiconductor electronics has overcome many technological obstacles, despite skeptics predicting the end of Moore's law due to these difficulties [2–6]. However, apart from these technological obstacles, there are fundamental limits for scaling from the laws of physics: For example, there are limits imposed by quantum

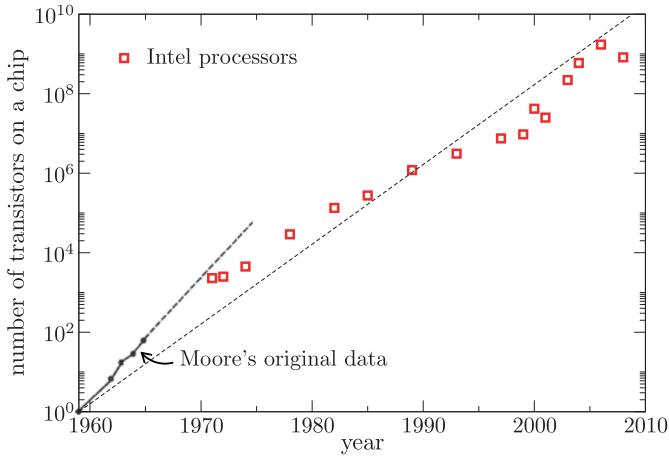


Figure 1.1 – Moore’s law on the example of Intel processors (data from www.intel.com/technology/mooreslaw/index.htm), together with Moore’s original prediction (from [1]). The long dashed line is a fit to an exponential, doubling every 18 months.

mechanics, such as the Heisenberg uncertainty principle [7, 8], or limits imposed by the material, such as heat removal or fluctuations in the number of dopants per transistor [9]. Today, it seems that the scaling in conventional semiconductor electronics may very well continue for about another two decades [9, 10]—but not beyond. In the meantime, the search for alternative approaches to a future electronics has already begun.

Whereas conventional electronics makes use of the *charge* of the electron, the field of *spintronics* seeks to use the electron *spin*. Spintronics is a wide field including topics with many different aspects (for reviews, see [11–13]). Amongst the earliest examples of spintronics are the various *magnetoresistance* effects, where the giant magnetoresistance (GMR) effect [14, 15] and the tunneling magnetoresistance TMR effect [16–18] are the most prominent examples. In both cases, the resistance of two (or more) ferromagnetic layers, separated by a spacer layer, depends on the relative angle between the magnetizations in the ferromagnets. While this spacer is a non-magnetic metal in the case of the GMR effect, and the underlying physics is governed by diffusive transport, it is an insulator in the case of the TMR effect, where the physics is governed by quantum mechanical tunneling. Apart from these differences, the magnetoresistance effect can be understood in both cases from the different density of states for spin up and down in the ferromagnets. Since the resistance depends strongly on the orientation of the layers, the GMR effect can be used as a very sensitive sensor for magnetic fields and has found its way into the read heads of hard drives, being a tremendous technological success. As a consequence, the 2007 Nobel Prize has been awarded to the discovery of the GMR effect. The most prominent application of the TMR is magnetic

random access memory (for a review of the applications of the GMR and TMR effect, see Ref. [19]).

The magnetoresistance effects are mostly applied to store and retrieve information. In order to use spin for logic operations, it must be possible to store, transport and manipulate the electron spin. Semiconductors are good candidates, as they typically exhibit a long spin relaxation time and allow the manipulation of spin via spin-orbit coupling, hence forming the field of *semiconductor spintronics* [13].

The spin field-effect transistor proposed in the seminal work of Datta and Das [20] is a good example for the fundamental ideas in a spin-based logic. In the spin field-effect transistor, source and drain contacts are formed by ferromagnetic metal contacts with parallel magnetizations. The electrons injected from the source pass through a two-dimensional electron gas, where the orientation of the spin can be manipulated by a gate voltage, tuning the Bychkov-Rashba spin-orbit coupling [21, 22]. When the orientation of the spin is not changed, it can easily exit through the drain contact (which has the same magnetization direction as the source contact) leading to a low resistance ("ON"-state), whereas it will be scattered back, if the spin has been flipped, leading to a large resistance ("OFF"-state). Switching from "ON" to "OFF" then only requires the energy needed to flip a spin, which is much smaller than the energy needed to remove an electron from the conduction channel in a conventional field-effect transistor. Hence, spintronics may lead to a logic with a reduced power consumption. Whilst it is not clear if the spin field-effect transistor of Datta and Das will ever be realized experimentally [13], a number of spin field-effect transistors based on similar ideas have been proposed, such as in Refs. [23, 24].

Another prominent spintronics example is the spin Hall effect, where in a charge current flow spin up and down are deflected into opposite directions due to spin-orbit coupling. The spin Hall effect comes in an extrinsic version [25, 26], where this spin-dependent deflection is due to scattering from impurities, and in an intrinsic version [27, 28] even in the absence of impurities, due to the spin-orbit coupling in semiconductors. The spin Hall effect may lead to a dissipationless, pure spin current [27], i.e. a spin current without any accompanying charge current, possibly paving the way towards dissipationless logic circuits.

The physics of graphene is another field that has seen a tremendous amount of interest since the experimental discovery of single layer graphene in 2004 [29]. Graphene is a rather remarkable material: It is a two-dimensional honeycomb lattice of carbon atoms, i.e. a single layer of graphite, and was believed not to exist, as two-dimensional crystals are thermodynamically unstable [30]. The fact that graphene nevertheless is found in the experiment, has been explained by a stabilization through the underlying substrate or intrinsic rippling [31].

In fact, now it is believed that every pencil trace contains a few flakes of single-layer graphene—the main problem is to find and identify them (for a review on the experimental fabrication and identification of graphene, see Ref. [31]). The low-energy physics of graphene is governed by the Dirac equation, and hence graphene exhibits many unique electronic properties, such as the odd-integer quantum Hall effect [32]

that can be observed even at room temperature [33] and Klein tunneling [34] (for reviews, see Refs. [31, 35]). Moreover, graphene shows an exceptionally high mobility [36, 37] even at room-temperature [29, 32]. Thus, with graphene it might be possible to realize a room-temperature ballistic field-effect transistor, allowing for operation at much higher frequencies than conventional transistors.

Apart from these unique electronic properties, graphene is also a promising candidate for spintronics and spin-based quantum computing [38], as the spin relaxation time in graphene is expected to be very long [39–41]. Spin injection from ferromagnetic metals has already been demonstrated experimentally [42–46].

In this thesis, we discuss two different spintronics examples: First, we study the magnetic field dependence of the tunneling anisotropic magnetoresistance (TAMR) effect, explaining recent experimental results [47, 48]. The TAMR effect is a prototypical example of a spintronics device, as it exhibits characteristics both of the magnetoresistive effects and of semiconductor spin-orbit coupling.

Second, we study the generation of spin currents in graphene-only devices. In particular, we show that the edge state in zigzag graphene nanoribbons can be used to generate spin-polarized currents. Moreover, this edge state can induce a spin Hall effect analogous to semiconductors with spin-orbit coupling, allowing for the generation of pure spin currents. The effects proposed in this work may help towards achieving an all-graphene based spintronics.

1.2 The need for numerics

In this work, we are concerned with calculating transport properties of systems that are neither very small, nor very large. For example, the graphene nanoribbons that we discuss in this work typically have a few ten to hundred thousand atoms. Hence, these systems are more complicated than a very small system where only the properties of a single atom may matter, but due to their finite size also more complicated than a very large system which may be approximated by bulk properties. When calculating the properties of such a system, one can choose between two approaches: Designing a simplified model, that can still be solved analytically, or performing numerical simulations on a more detailed description of the system.

Numerical simulations are a crucial component whenever more quantitative predictions are sought. In addition, numerics is a valuable tool for validating the predictions of simplified models. Moreover, numerical simulations can give an insight into the physical processes of not yet understood, complicated systems. From the information gained in the numerics it is then often possible to build a simplified model, capturing the essential physics. In this work we are concerned with advanced numerical calculations of spin transport properties which also serve as reference calculations for corresponding analytical models.

In fact, numerics has been in the toolbox of physicists even before the invention of electronic computers. The first large-scale numerical calculation seems to have been

carried out by Alexis-Claude Clairaut in 1758, for computing the orbit of comet Halley. Working together with two friends, the numerical calculations took nearly five months [49]. Of course, the advent of electronic computers has changed the way numerical computations are done. As computers have become more and more powerful, numerical simulations can deal with ever more complex problems. However, these increased capabilities have also led to scientific computer programs that become more and more complex. When writing a code for numerical computations, it is now also important to follow general programming paradigms.

One such paradigm is *code reusability*. Code reusability means that program code should not be written specifically for a single problem, but should be applicable to a general class of problems. Applying such a general code decreases the necessary development time. Moreover, any time code is written, mistakes are made, and hence reusing existing code can help avoiding mistakes. In addition, code reusability can also help to increase code quality, as the respective program parts can be tested thoroughly.

From the point of view of physics, the development of *generic* algorithms, i.e. algorithms that can be applied to a wide class of physical problems, is a prerequisite for code reusability. In this work, we will develop new algorithms for a generic approach to quantum transport, applicable to *arbitrary* tight-binding systems. These techniques serve as the foundation of the numerical simulations of this work. Moreover due to its generic nature, the program developed in the course of this work has also been applied to other transport problems [50–57].

1.3 Outline

Every chapter of this thesis starts with an introductory section, motivating and/or defining the problem under investigation, and ends with a summary of the main findings.

The thesis is organized in two parts: In the first part we develop new algorithms for a generic approach to transport in arbitrary tight-binding models.

To this end, **Chapter 2** reviews the established transport formalism based on Green's functions and the scattering approach of Landauer and Büttiker. In particular, the equivalence of the Green's function and the scattering approach is emphasized. This equivalence can be written in a compact form using the generalized Fisher-Lee relation of Appendix C. From the general transport theory, we identify two challenges on the way to a generic and efficient numerical approach for transport in tight-binding models: the calculation of the lead Green's function and the efficient calculation of the retarded Green's function of the system. Novel solutions to these problems are then presented in the following two chapters.

Chapter 3 is concerned with the calculation of the Green's function of the leads. For this we prove rigorously a general expression for the lead Green's function, extending the expressions known from previous works. Since this general expression is found to be numerically unstable in certain, important systems, we then develop a new, numerically

stable algorithm to evaluate the Green's function of the lead.

In **Chapter 4** we develop a matrix reordering algorithm based on graph partitioning techniques that brings the Hamiltonian matrix of a tight-binding system to a form optimal for transport. In particular, through this reordering well-established quantum transport algorithms that were previously restricted to linear geometries can now be applied to arbitrary geometries, including multi-terminal structures. Moreover, the matrix reordering can lead to a significant speed-up of calculations.

The results of the first part together with Appendix C are the foundation of a general transport code that can be applied to arbitrary tight-binding systems. In the second part of this thesis, this transport code is used to perform numerical simulations of spintronics effects in magnetic tunnel junctions and graphene.

In **Chapter 5** we investigate the magnetic field dependence of the TAMR effect in an epitaxially grown Fe/GaAs/Au tunnel junction, in order to explain recent experimental findings. To this end, we extend a previously developed model that relates the TAMR effect to the effects of spin-orbit coupling in the tunnel barrier to include the orbital effects of the magnetic fields. The characteristic features of the experimental findings are reproduced by the numerical simulations, and the underlying physics is highlighted within a qualitative model.

Chapter 6 is devoted to an extensive study of the charge transport properties of the graphene edge state. The graphene edge state is a peculiar state localized at a zigzag edge of graphene. In particular, we find that the paradigm model of transport in graphene, the nearest-neighbor tight-binding approximation, cannot be applied to study transport properties of the edge state, as exponentially small corrections to this model can change the edge state properties fundamentally. This is demonstrated by numerical simulations and can be understood from symmetry considerations in a perturbative analysis of the edge state. We especially highlight the importance of next-nearest neighbor hopping for the graphene edge state.

These results on the *charge* transport properties of the edge state are the foundation for the *spin* transport properties of graphene nanoribbons, studied in **Chapter 7**. We show ways how to generate spin-polarized and pure spin currents from the theoretically predicted edge magnetism. In particular, rough graphene nanoribbons are found to be a natural source of spin-polarized electrons, exhibiting universal spin conductance fluctuations. Moreover, the edge state leads to a geometrically induced spin Hall effect, that can be used to generate spin-polarized currents in three-terminal devices, and pure spin currents in four-terminal devices. The results of this chapter show an alternative to ferromagnetic metal contacts for generating spin currents, paving the way to an all-graphene based spintronics.

Finally, the results of this thesis are summarized and further perspectives are discussed.

The appendix contains additional information and technical details complementing the main text. **Appendix A** gives general expressions for probability and current densities in tight-binding models, including spin. **Appendix B** reviews the recursive Green's function method that is the quantum transport algorithms used in this work,

including its extension to non-equilibrium. In **Appendix C** we derive a novel, compact form of a generalized Fisher-Lee relation, valid for any tight-binding model. This Fisher-Lee relation connects transmission and reflection amplitudes with properties of the Green's function, thus bringing together the Green's function and the scattering formalism. In **Appendix D** we present the details of the derivations of some equations used in Chapter 3. **Appendix E** reviews in more detail the Fiduccia-Mattheyses graph partitioning algorithm employed in Chapter 4. In **Appendix F** we review the method of finite differences, used in Chapter 5, that recasts the Schrödinger equation into an effective tight-binding problem. Finally, **Appendix G** reviews in detail the tight-binding model for graphene and the derivation of the effective, low-energy Dirac equation. Both approaches to graphene are employed extensively in Chapters 6 and 7.

Part I

Computational concepts: A generic
approach to transport in
tight-binding models

“Seit Jahrhunderten hatten die Astronomen Verfahren zur Berechnung der Bewegung eines Himmelskörpers entwickelt (...) Die Mitarbeiter des Astronomischen Recheninstituts, das damals in Berlin-Babelsberg angesiedelt war (...), hatten viel Erfahrung in der Berechnung von Planetenephemeriden. Ein Jahr vor der Annäherung von Amor an die Erde [Author’s note: this was in 1955] schritt man ans Werk. Doch damals arbeitete in Göttingen bereits die G2 [Author’s note: an early computer], und so bot es sich an, auch die Astrophysiker des dortigen Max-Planck-Instituts einzuladen, die Bahn des Amor zu berechnen. Da würde man sehen, was die neuen Maschinen, von denen so viel die Rede war, wirklich zu leisten vermochten. (...)

Doch das Ergebnis der Göttinger unterschied sich deutlich von den in Babelsberg nach klassischen Verfahren berechneten Ephemeriden. (...) Das war alles andere als die Empfehlung, in Zukunft solche Rechnungen einem Computer zu überlassen.

Und dann kam Amor. Wo aber stand er am Himmel? Dort wo ihm die Babelsberger Rechnungen seinen Platz zugewiesen haben? Nein, er stand genau dort, wohin ihn die G2 platziert hatte! Im Jahr darauf konnte man im Jahresbericht des Babelsberger Instituts lesen: »Soweit ein Urteil schon jetzt möglich ist, ist bei Rechnungen der höchsten Genauigkeit die Maschine der Handrechnung überlegen.«

Heute ist das eine Binsenweisheit, aber auch Binsenweisheiten setzen sich eben nicht immer leicht durch”

Rudolf Kippenhahn – *Amor und der Abstand zur Sonne*

Chapter 2

Green's function formalism for transport

2.1 Introduction

Calculating transport properties of a system involves a complicated many-body problem that as such cannot be hoped to be solved exactly. Therefore, any transport calculation involves approximations motivated by physical arguments. Here, we will consider systems with a considerable number of electrons, i.e. high density, and strong coupling to external leads. In this limit, the system can be modeled usually in terms of non-interacting particles in an effective potential, resulting from external and mean-field potentials [58]. It will not be the subject of this work to derive such an effective Hamiltonian from first principles, but we will instead use well-established models in order to derive new physical phenomena.

A common approach to transport is based on the Green's function formalism [58–63]. In particular, the non-equilibrium Green's function (NEGF) formalism pioneered by Kadanoff and Baym [64] and Keldysh [65] has gained a lot of popularity in the recent years. The non-equilibrium Green's function formalism is capable to describe systems including a finite bias and interactions. In addition, the central results of the NEGF formalism can be summarized in a few equations that can be conveniently implemented on a computer. Due to this fact, the NEGF formalism is more and more applied also to non-interacting systems, and so will we do in this work.

However, the generality and power of the non-equilibrium Green's function formalism comes at the cost of complicating the physical understanding, as an intuitive picture of the underlying physics is not obvious there, and a rather involved derivation of the central results. Especially for the beginner, the NEGF formalism is thus hard to understand. Being a *formalism* it can still be applied to a given problem without

full understanding, but the results obtained thereby lack physical intuition. In fact, an intuitive picture may not be possible for an interacting system. However for non-interacting systems, such an intuitive understanding is still possible, as we discuss below. Applying the non-equilibrium Green's function formalism to non-interacting systems is justified because of the numerical convenience it provides, but may comprise the danger of obscuring the physical intuition, if not discussed properly.

Below, we will therefore discuss the non-equilibrium Green's function formalism in detail. Being an introduction to the subject, this chapter tries to be as self-contained as possible. We start with some basic definitions and introduce the concept of Green's functions. In order to give a physical meaning to these Green's functions, we will explicitly calculate them for the case of a non-interacting system. We then derive the main results of the non-equilibrium Green's function formalism in the form of a perturbation theory. From these results we develop a set of transport equations in tight-binding approximation, suitable for a computer implementation. By showing connections between the non-equilibrium Green's function and the scattering formalism, we will then provide an intuitive picture of transport valid for non-interacting systems, that to some extent also carries over to interacting systems. We conclude this chapter by identifying problems that remain to be solved in order to apply the non-equilibrium formalism to arbitrary systems. Solutions to these problems will then be presented in the following two chapters.

2.2 Basic definitions

We start with a very general description of a many-body system with the Hamiltonian

$$\mathcal{H}(t) = \mathcal{H}_0 + \mathcal{H}_{\text{int}} + \mathcal{H}_{\text{ext}}(t) = \mathcal{H}_0 + \mathcal{H}'(t), \quad (2.1)$$

where \mathcal{H}_0 is the Hamiltonian of a non-interacting system, \mathcal{H}_{int} describes interactions and $\mathcal{H}_{\text{ext}}(t)$ is a (possibly) time-dependent external perturbation which is assumed to vanish before some time t_0 , $\mathcal{H}_{\text{ext}}(t) = 0$ for $t < t_0$. The time-dependence of a quantum mechanical (many-body) state $|\psi\rangle_S$ is then governed by the Schrödinger equation,

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle_S = \mathcal{H}(t) |\psi(t)\rangle_S . \quad (2.2)$$

Note that we use here the Schrödinger picture of quantum mechanics, where all the time-dependence is carried by the quantum mechanical state. Since any theory of Green's functions in many-body systems inevitably involves using different quantum mechanical pictures, we will carefully distinguish these by appropriate subscripts whenever such a distinction is necessary.

The Schrödinger equation can be formally solved by some time evolution operator $U(t, t_0)$, that evolves the state from time t_0 to t ,

$$|\psi(t)\rangle_S = U(t, t_0) |\psi(t_0)\rangle_S . \quad (2.3)$$

The time evolution operator itself is formally obtained by integrating the Schrödinger equation,

$$i\hbar \frac{\partial}{\partial t} U(t, t_0) = \mathcal{H}(t) U(t, t_0), \quad (2.4)$$

as

$$U(t, t_0) = \mathcal{T} e^{-\frac{i}{\hbar} \int_{t_0}^t dt' \mathcal{H}(t')}, \quad (2.5)$$

where \mathcal{T} is the time ordering operator that orders operators according to their time argument, with the latest times to the left.

The time-dependent expectation value of an observable A with respect to some pure quantum mechanical state is given as

$$\bar{A}(t) = \langle \psi(t) |_S A_S | \psi(t) \rangle_S, \quad (2.6)$$

where A_S is an operator in the Schrödinger picture. However, commonly we will consider the more general situation that the system is not in a pure state, but in a statistical mixture, such that

$$\bar{A}(t) = \sum_i p_i \langle \psi_i(t) |_S A_S | \psi_i(t) \rangle_S, \quad (2.7)$$

where p_i is the probability of being in state ψ_i . Introducing the density matrix $\rho_S(t) = \sum_i p_i |\psi_i(t)\rangle_S \langle \psi_i(t)|_S$, this expectation value can be written in a compact form,

$$\bar{A}(t) = \text{Tr}(\rho_S(t) A_S). \quad (2.8)$$

Note that the density matrix $\rho_S(t)$ is not an operator in the usual sense—contrary to the usual Schrödinger operators it is time-dependent. This is due to the time-dependence of the quantum mechanical states, and the density matrix obeys the von Neumann equation

$$i\hbar \frac{\partial}{\partial t} \rho_S(t) = [H(t), \rho_S(t)] \quad (2.9)$$

as can be seen easily from the Schrödinger equation (2.2). Here, $[A, B] = AB - BA$ is the commutator.

Contrary to the Schrödinger picture, in the Heisenberg picture all the time dependence is carried by the operators, whereas the states are time-independent. The Schrödinger and the Heisenberg picture are supposed to coincide at some time t_0 , such that

$$|\psi\rangle_{\mathcal{H}} = |\psi(t_0)\rangle_S \text{ and } A_{\mathcal{H}}(t_0) = A_S. \quad (2.10)$$

Operators in the Heisenberg picture are then related to operators in the Schrödinger picture via the time evolution operator $U(t, t_0)$, Eq. (2.5),

$$A_{\mathcal{H}}(t) = U^\dagger(t, t_0) A_S U(t, t_0), \quad (2.11)$$

and they obey the equation of motion¹

$$i\hbar \frac{\partial}{\partial t} A_{\mathcal{H}}(t) = [A_{\mathcal{H}}(t), \mathcal{H}_{\mathcal{H}}(t)] . \quad (2.12)$$

Since the Heisenberg states are time-independent, so is the density matrix $\rho_{\mathcal{H}}$, and the expectation value of an observable is given as

$$\bar{A}(t) = \text{Tr}(\rho_{\mathcal{H}} A_{\mathcal{H}}(t)) . \quad (2.13)$$

In this work, we will assume that the system is initially in thermal equilibrium, for $t < t_0$. In order to simplify the discussion, we take the limit $t_0 \rightarrow -\infty$ and assume that the interaction Hamiltonian \mathcal{H}_{int} is switched off adiabatically when going back in time. Then the grand canonical density matrix depends only on \mathcal{H}_0 ,²

$$\rho_{\mathcal{H}} = \frac{e^{-\beta(\mathcal{H}_0 - \mathcal{Y}_0)}}{\text{Tr}(e^{-\beta(\mathcal{H}_0 - \mathcal{Y}_0)})} = \frac{1}{Z} e^{-\beta(\mathcal{H}_0 - \mathcal{Y}_0)} . \quad (2.14)$$

Here, $\beta = 1/k_B T$, where T is the temperature, k_B the Boltzmann constant, and $Z = \text{Tr}(e^{-\beta(\mathcal{H}_0 - \mathcal{Y}_0)})$ the partition function. In general, we will consider systems consisting of parts with different electrochemical potential. Consequently, we assume \mathcal{Y}_0 to be of the general form $\mathcal{Y}_0 = \sum_i \mu_i \mathcal{N}_i$, where μ_i is the electrochemical potential and \mathcal{N}_i the number operator for the respective subsystem [69]. A particular example is a system consisting of leads with different electrochemical potentials, as considered in Section 2.5.2. With respect to the grand canonical ensemble, the expectation value of an observable is given as

$$\bar{A}(t) = \langle A_{\mathcal{H}}(t) \rangle = \frac{\text{Tr}(e^{-\beta(\mathcal{H}_0 - \mathcal{Y}_0)} A_{\mathcal{H}}(t))}{\text{Tr}(e^{-\beta(\mathcal{H}_0 - \mathcal{Y}_0)})} . \quad (2.15)$$

Many-body operators are most conveniently written in terms of creation and annihilation operators in second quantization (for an introduction see, e.g. [70]). The

¹ Note that the Hamiltonian in the Heisenberg picture is defined as $\mathcal{H}_{\mathcal{H}}(t) = U^\dagger(t, t_0) \mathcal{H}(t) U(t, t_0)$. If \mathcal{H} is time-independent, we have $\mathcal{H}_{\mathcal{H}}(t) = \mathcal{H}$, as then $U(t, t_0) = e^{i\frac{\hbar}{\text{E}}\mathcal{H}(t-t_0)}$ commutes with \mathcal{H} . However, if $\mathcal{H}(t)$ is explicitly time-dependent, $U(t, t_0)$ does not necessarily commute with $\mathcal{H}(t)$, as $\mathcal{H}(t)$ itself does not necessarily commute with $\mathcal{H}(t')$ at some other time $t' \neq t$.

² In taking this limit, we neglect initial correlations imposed by \mathcal{H}_{int} . It is in principle possible to include these correlations when developing a perturbation theory for the Green's function, as is, for example, done in Refs. [61, 63, 64, 66]. These accounts take the limit $t_0 \rightarrow -\infty$ *after* the perturbation theory has been developed—but this is equivalent to already taking the limit $t_0 \rightarrow -\infty$ in the very beginning, as we do here. Taking the limit early simplifies the subsequent discussion considerably, and nevertheless yields identical results.

Usually it is argued that initial correlations are washed out by the interactions, as eventually steady state is reached [61]. It is possible to develop a perturbation theory *including* the initial correlations [67, 68], but such a treatment becomes quite involved. In any case, as we are not considering interacting systems in this work anyway, we need not worry about this point.

operator c_n (c_n^\dagger) removes (creates) a particle in state $|n\rangle$, with corresponding single-particle wave function $\varphi_n(\mathbf{x})$. Fermion statistics is imposed by the anti-commutation relations

$$\{c_n^\dagger, c_m\} = \delta_{nm}, \quad (2.16a)$$

$$\{c_n^\dagger, c_m^\dagger\} = \{c_n, c_m\} = 0, \quad (2.16b)$$

where $\{A, B\} = AB + BA$ is the anti-commutator. If the $\varphi_n(\mathbf{x})$ form a complete set, we can define *field operators* as

$$\begin{aligned} \psi(\mathbf{x}) &= \sum_n \varphi_n(\mathbf{x}) c_n, \text{ and} \\ \psi^\dagger(\mathbf{x}) &= \sum_n \varphi_n^*(\mathbf{x}) c_n^\dagger, \end{aligned} \quad (2.17)$$

that remove and add a particle at point \mathbf{x} , respectively. The field operators obey the anti-commutation relations

$$\{\psi(\mathbf{x})^\dagger, \psi(\mathbf{x}')\} = \delta(\mathbf{x} - \mathbf{x}'), \quad (2.18a)$$

$$\{\psi(\mathbf{x})^\dagger, \psi^\dagger(\mathbf{x}')\} = \{\psi(\mathbf{x}), \psi(\mathbf{x}')\} = 0. \quad (2.18b)$$

These commutation relations are valid for both the Schrödinger and the Heisenberg picture. It should be noted though that in the Heisenberg picture both operators have to be evaluated at the same time t . Otherwise, the complicated time-evolution leads to non-trivial relations.

Green's functions are defined as an expectation value of two or more field operators. There is a number of different definitions for Green's functions, each of them useful in a specific situation. In this work, we will consider in particular the retarded, the advanced, the lesser and the greater Green's function:

$$G^r(\mathbf{x}, t, \mathbf{x}', t') = -\frac{i}{\hbar} \Theta(t - t') \langle \{\psi_{\mathcal{H}}(\mathbf{x}, t), \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t')\} \rangle, \quad (2.19a)$$

$$G^a(\mathbf{x}, t, \mathbf{x}', t') = \frac{i}{\hbar} \Theta(t' - t) \langle \{\psi_{\mathcal{H}}(\mathbf{x}, t), \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t')\} \rangle, \quad (2.19b)$$

$$G^<(\mathbf{x}, t, \mathbf{x}', t') = \frac{i}{\hbar} \langle \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t') \psi_{\mathcal{H}}(\mathbf{x}, t) \rangle, \quad (2.19c)$$

$$G^>(\mathbf{x}, t, \mathbf{x}', t') = -\frac{i}{\hbar} \langle \psi_{\mathcal{H}}(\mathbf{x}, t) \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t') \rangle, \quad (2.19d)$$

where $\Theta(t)$ is the Heaviside step function. The anti-commutator in the definitions of advanced and retarded Green's function involves field operators at different times and can thus not be evaluated with the anti-commutation relations at equal times, Eq. (2.18a).

These various Green's functions carry different information. For example, the lesser Green's function directly allows the evaluation of physical observables, such as the local charge carrier density,

$$n(\mathbf{x}, t) = \langle \psi_{\mathcal{H}}^\dagger(\mathbf{x}, t) \psi_{\mathcal{H}}(\mathbf{x}, t) \rangle = -i\hbar G^<(\mathbf{x}, t, \mathbf{x}, t). \quad (2.20)$$

As a further example, the retarded Green's function can be used to extract the transmission probability of charge carriers through a system (see Section 2.5.2).

From the definitions (2.19a)–(2.19d) it is obvious that these various Green's functions are not independent. For example, we find the useful identities

$$G^r(\mathbf{x}, t, \mathbf{x}', t') = (G^a(\mathbf{x}', t', \mathbf{x}, t))^*, \quad (2.21)$$

$$G^<(\mathbf{x}, t, \mathbf{x}', t') = -(G^<(\mathbf{x}', t', \mathbf{x}, t))^*, \quad (2.22)$$

$$G^r - G^a = G^> - G^<. \quad (2.23)$$

In equilibrium [58] and in non-equilibrium steady state [61] the Green's functions only depend on the time difference, $G(\mathbf{x}, t, \mathbf{x}', t') = G(\mathbf{x}, \mathbf{x}'; t - t')$. In this situation, it is useful to introduce the energy-dependent Green's function through Fourier transformation,

$$G(\mathbf{x}, \mathbf{x}'; E) = \int_{-\infty}^{\infty} dt e^{\frac{i}{\hbar} Et} G(\mathbf{x}, \mathbf{x}'; t). \quad (2.24)$$

The inverse Fourier transformation is then given by

$$G(\mathbf{x}, \mathbf{x}'; t - t') = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dE e^{-\frac{i}{\hbar} E(t-t')} G(\mathbf{x}, \mathbf{x}'; E). \quad (2.25)$$

For the purpose of developing a perturbation theory for the ground state, the causal (or time-ordered) Green's function

$$G^c(\mathbf{x}, t, \mathbf{x}', t') = -\frac{i}{\hbar} \langle \mathcal{T} \psi_{\mathcal{H}}(\mathbf{x}, t) \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t') \rangle \quad (2.26)$$

plays a crucial role (see, e.g. [58]). Here, \mathcal{T} denotes the time ordering operator for fermionic operators,

$$\mathcal{T} A(t_1) B(t_2) = \begin{cases} +A(t_1) B(t_2) & \text{for } t_1 > t_2, \\ -B(t_2) A(t_1) & \text{for } t_2 > t_1, \end{cases} \quad (2.27)$$

and analogously for more than two operators, with a minus sign for every operator interchange³. A generalization of this causal Green's function will be the foundation of a non-equilibrium perturbation theory in Section 2.4. However, before rising to this challenge, we first examine the various Green's functions defined here in the limit of a non-interacting system.

³ There was no minus sign in the definition of the time ordering operator in Eq. (2.5): The operator to be ordered there, $\mathcal{H}(t)$, always consists of an even number of fermionic operators, as does any observable.

2.3 Green's functions of non-interacting systems

We now consider the particular case of a non-interacting system, where $\mathcal{H} = \mathcal{H}_0$. The Hamiltonian can then be written as

$$\mathcal{H}_0 = \int d\mathbf{x} \psi^\dagger(\mathbf{x}) H_0(\mathbf{x}) \psi(\mathbf{x}), \quad (2.28)$$

where $H_0(\mathbf{x})$ is a single particle Hamiltonian⁴. Alternatively, introducing the set of eigenstates $\{\varphi_n\}$ of the single particle Hamiltonian, $H_0\varphi_n = E_n\varphi_n$, we can write the Hamiltonian as

$$\mathcal{H}_0 = \sum_n E_n c_n^\dagger c_n, \quad (2.29)$$

where c_n (c_n^\dagger) removes (creates) a particle in state φ_n .

Since the non-interacting Hamiltonian is time-independent, the time-dependence of the field operator in the Heisenberg picture is given by

$$\begin{aligned} \psi_{\mathcal{H}_0}(\mathbf{x}, t) &= e^{\frac{i}{\hbar}\mathcal{H}_0(t-t_0)} \psi_{\mathcal{H}_0}(\mathbf{x}, t_0) e^{-\frac{i}{\hbar}\mathcal{H}_0(t-t_0)} \\ &= e^{\frac{i}{\hbar}\mathcal{H}_0(t-t_0)} \psi(\mathbf{x}) e^{-\frac{i}{\hbar}\mathcal{H}_0(t-t_0)}. \end{aligned} \quad (2.30)$$

Here, we use the subscript \mathcal{H}_0 to emphasize that the dynamics is governed by the non-interacting Hamiltonian. The equation of motion for the field operator is then

$$i\hbar \frac{\partial}{\partial t} \psi_{\mathcal{H}_0}(\mathbf{x}, t) = [\psi_{\mathcal{H}_0}(\mathbf{x}, t), \mathcal{H}_0] = H_0(\mathbf{x}) \psi_{\mathcal{H}_0}(\mathbf{x}, t), \quad (2.31)$$

where we made use of the anti-commutation relations (2.18). In the same manner, we find

$$i\hbar \frac{\partial}{\partial t} c_{n,\mathcal{H}_0}(t) = [c_{n,\mathcal{H}_0}(\mathbf{x}, t), \mathcal{H}_0] = E_n c_{n,\mathcal{H}_0}(t) \quad (2.32)$$

that can be easily integrated as

$$c_{n,\mathcal{H}_0}(t) = e^{-\frac{i}{\hbar}E_n(t-t_0)} c_n. \quad (2.33)$$

Thus, the time-dependence of the field operator can be explicitly written as

$$\psi_{\mathcal{H}_0}(\mathbf{x}, t) = \sum_n \varphi_n(\mathbf{x}) e^{-\frac{i}{\hbar}E_n(t-t_0)} c_n. \quad (2.34)$$

Equipped with these expressions for the time dependence of the field operators, we now turn to the Green's functions. Note that we subsequently mark these with a

⁴ For example, the Hamiltonian of an electron in a potential V , $H_0(\mathbf{x}) = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x})$

subscript as G_0 , in order to indicate the dependence on the non-interacting Hamiltonian \mathcal{H}_0 . We begin with deriving an equation of motion for the retarded Green's function:

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} G_0^r(\mathbf{x}, t, \mathbf{x}', t') &= i\hbar \frac{\partial}{\partial t} \left[-\frac{i}{\hbar} \Theta(t - t') \langle \{ \psi_{\mathcal{H}_0}(\mathbf{x}, t), \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') \} \rangle \right] \\ &= \delta(t - t') \langle \{ \psi_{\mathcal{H}_0}(\mathbf{x}, t), \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') \} \rangle \\ &\quad - \frac{i}{\hbar} \Theta(t - t') \langle \{ i\hbar \frac{\partial}{\partial t} \psi_{\mathcal{H}_0}(\mathbf{x}, t), \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') \} \rangle. \end{aligned} \quad (2.35)$$

In the first term we have set $t = t'$ because of the delta function. Using the anti-commutation relation (2.18a) to simplify the first term and the equation of motion for the field operator (2.31) for the second term, we find

$$\left(i\hbar \frac{\partial}{\partial t} - H_0(\mathbf{x}) \right) G_0^r(\mathbf{x}, t, \mathbf{x}', t') = \delta(t - t') \delta(\mathbf{x} - \mathbf{x}'). \quad (2.36)$$

The retarded Green's function thus obeys the single particle Schrödinger equation with a delta function source term⁵. In contrast, we find for the lesser Green's function:

$$\left(i\hbar \frac{\partial}{\partial t} - H_0(\mathbf{x}) \right) G_0^<(\mathbf{x}, t, \mathbf{x}', t') = 0. \quad (2.37)$$

Equivalent equations hold for the advanced and greater Green's functions. Using the notation $G_0^{-1} = (i\hbar \frac{\partial}{\partial t} - H_0(\mathbf{x}))$, we can summarize these results formally as

$$G_0^{-1} G_0^{r,a} = \mathbb{1} \quad \text{and} \quad G_0^{-1} G_0^{<,>} = 0 \quad (2.38)$$

After deriving the equation of motions for the Green's functions for non-interacting systems, we now turn to explicitly calculating them.

The retarded Green's function is given as

$$\begin{aligned} G_0^r(\mathbf{x}, t, \mathbf{x}', t') &= -\frac{i}{\hbar} \Theta(t - t') \langle \{ \psi_{\mathcal{H}_0}(\mathbf{x}, t), \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') \} \rangle \\ &= -\frac{i}{\hbar} \Theta(t - t') \sum_{n,m} \varphi_n(\mathbf{x}) \varphi_m^*(\mathbf{x}') e^{-\frac{i}{\hbar}(E_n(t-t_0)-E_m(t'-t_0))} \underbrace{\langle \{ c_n, c_m^\dagger \} \rangle}_{=\delta_{n,m}} \\ &= -\frac{i}{\hbar} \Theta(t - t') \sum_n \varphi_n(\mathbf{x}) \varphi_n^*(\mathbf{x}') e^{-\frac{i}{\hbar} E_n(t-t')}, \end{aligned} \quad (2.39)$$

where we made use of Eq. (2.34) and the anti-commutation relations (2.16). Note that the dependence on the density matrix dropped out completely.

⁵ The Green's function thus obeys a equation of the form $LG(x, x') = \delta(x - x')$, where L is a linear differential operator. This kind of equation is known in the context of inhomogeneous differential equations, e.g. in electrodynamics, where the concept of a Green's function was introduced first. This similarity in the the equation of motion is the reason for the name "Green's function" in the context discussed here.

In order to calculate the energy dependent retarded Green's function, we use the representation

$$\Theta(t) = -\frac{1}{2\pi i} \int_{-\infty}^{\infty} dE' e^{-\frac{i}{\hbar} E' t} \frac{1}{E' + i\eta} \quad (2.40)$$

for the Heaviside step function, where η is an infinitesimally small positive number. We arrive at

$$\begin{aligned} G_0^r(\mathbf{x}, \mathbf{x}'; E) &= \int_{-\infty}^{\infty} dt e^{\frac{i}{\hbar} Et} G_0^r(\mathbf{x}, \mathbf{x}'; t) \\ &= \frac{1}{2\pi\hbar} \sum_n \varphi_n(\mathbf{x}) \varphi_n^*(\mathbf{x}') \int_{-\infty}^{\infty} dE' \int_{-\infty}^{\infty} dt e^{\frac{i}{\hbar}(E-E_n-E')t} \frac{1}{E' + i\eta} \\ &= \sum_n \varphi_n(\mathbf{x}) \varphi_n^*(\mathbf{x}') \int_{-\infty}^{\infty} dE' \delta(E - E_n - E') \frac{1}{E' + i\eta} \\ &= \sum_n \frac{\varphi_n(\mathbf{x}) \varphi_n^*(\mathbf{x}')}{E - E_n + i\eta}. \end{aligned} \quad (2.41)$$

The non-interacting many-body retarded Green's function is thus identical to the well-known single-particle retarded Green's function [59, 60]. The retarded Green's function does not carry any information about temperature or chemical potential, it only depends on the eigenfunctions and -energies of the single-particle Hamiltonian H_0 . From Eq. (2.21) we immediately find an expression for the advanced Green's function,

$$G_0^a(\mathbf{x}, \mathbf{x}'; E) = \sum_n \frac{\varphi_n(\mathbf{x}) \varphi_n^*(\mathbf{x}')}{E - E_n - i\eta}. \quad (2.42)$$

These expressions may be written in a more compact form in terms of an operator identity, using $H_0 |\varphi_n\rangle = E_n |\varphi_n\rangle$,

$$\begin{aligned} G_0^{r(a)}(\mathbf{x}, \mathbf{x}'; E) &= \langle \mathbf{x} | \sum_n \frac{|\varphi_n\rangle \langle \varphi_n|}{E - E_n \pm i\eta} |\mathbf{x}'\rangle \\ &= \langle \mathbf{x} | \left(\frac{1}{E - H_0 \pm i\eta} \underbrace{\sum_n |\varphi_n\rangle \langle \varphi_n|}_{=1} \right) |\mathbf{x}'\rangle \\ &= \langle \mathbf{x} | (E - H_0 \pm i\eta)^{-1} |\mathbf{x}'\rangle. \end{aligned} \quad (2.43)$$

In order to calculate $G_0^<$ and $G_0^>$, we derive relations connecting them with G_0^r and G_0^a , following an argument by Kadanoff and Baym [64]. We assume that the system is in equilibrium, governed by a single electrochemical potential μ . Then, the density matrix is of the form $e^{-\beta(\mathcal{H}_0 - \mu\mathcal{N})}$ and we find

$$\begin{aligned} G_0^<(\mathbf{x}, t, \mathbf{x}', 0) &= \frac{i}{\hbar} \frac{1}{Z} \text{Tr} \left(e^{-\beta(\mathcal{H}_0 - \mu\mathcal{N})} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', 0) \psi_{\mathcal{H}_0}(\mathbf{x}, t) \right) \\ &= \frac{i}{\hbar} \frac{1}{Z} \text{Tr} \left(e^{-\beta(\mathcal{H}_0 - \mu\mathcal{N})} \psi_{\mathcal{H}_0}(\mathbf{x}, t) e^{-\beta(\mathcal{H}_0 - \mu\mathcal{N})} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', 0) e^{+\beta(\mathcal{H}_0 - \mu\mathcal{N})} \right), \end{aligned} \quad (2.44)$$

where we made use of the cyclic invariance of the trace. Since $\psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', 0)$ creates a particle, we have

$$\psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', 0) f(\mathcal{N}) = f(\mathcal{N} - 1) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', 0) \quad (2.45)$$

for any function $f(\mathcal{N})$ of the number operator, as can be seen by operating on states with a given number of particles. Therefore, we find

$$e^{\beta\mu\mathcal{N}} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}, 0) e^{-\beta\mu\mathcal{N}} = e^{\beta\mu} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}, 0), \quad (2.46)$$

and thus

$$\begin{aligned} e^{-\beta(\mathcal{H}_0 - \mu\mathcal{N})} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}, 0) e^{\beta(\mathcal{H}_0 - \mu\mathcal{N})} &= e^{-\beta\mathcal{H}_0} e^{\beta\mu\mathcal{N}} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}, 0) e^{-\beta\mu\mathcal{N}} e^{\beta\mathcal{H}_0} \\ &= e^{\beta\mu} e^{-\beta\mathcal{H}_0} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}, 0) e^{\beta\mathcal{H}_0} \\ &= e^{\beta\mu} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}, i\hbar\beta). \end{aligned} \quad (2.47)$$

Here we used the fact that the number operator \mathcal{N} commutes with the Hamiltonian, and an analytic continuation of Eq. (2.30) into the complex plane. We finally arrive at

$$\begin{aligned} G_0^<(\mathbf{x}, t, \mathbf{x}', 0) &= \frac{i}{\hbar} e^{\beta\mu} \frac{1}{Z} \text{Tr} \left(e^{-\beta(\mathcal{H}_0 - \mu\mathcal{N})} \psi_{\mathcal{H}_0}(\mathbf{x}, t) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}, i\hbar\beta) \right) \\ &= -e^{\beta\mu} G_0^>(\mathbf{x}, t, \mathbf{x}', i\hbar\beta). \end{aligned} \quad (2.48)$$

Taking the Fourier transform, we find the energy-dependent lesser Green's function,

$$\begin{aligned} G_0^<(\mathbf{x}, \mathbf{x}'; E) &= -e^{\beta\mu} \int_{-\infty}^{\infty} dt e^{\frac{i}{\hbar} Et} G_0^>(\mathbf{x}, \mathbf{x}'; t - i\hbar\beta) \\ &= -e^{-\beta(E - \mu)} \int_{-\infty - i\hbar\beta}^{\infty - i\hbar\beta} dt e^{\frac{i}{\hbar} Et} G_0^>(\mathbf{x}, \mathbf{x}'; t) \\ &= -e^{-\beta(E - \mu)} G_0^>(\mathbf{x}, \mathbf{x}'; E), \end{aligned} \quad (2.49)$$

where we made use of the fact that $G_0^>(\mathbf{x}, \mathbf{x}'; t - t')$ is an analytic function for $-\hbar\beta < \Im(t - t') < 0$ [63, 64]. We now define the spectral density as

$$\begin{aligned} A(\mathbf{x}, \mathbf{x}'; E) &= i(G_0^r(\mathbf{x}, \mathbf{x}'; E) - G_0^a(\mathbf{x}, \mathbf{x}'; E)) \\ &= i(G_0^>(\mathbf{x}, \mathbf{x}'; E) - G_0^<(\mathbf{x}, \mathbf{x}'; E)), \end{aligned} \quad (2.50)$$

where the second equality is due to Eq. (2.23). Inserting Eq. (2.49) we find

$$A(\mathbf{x}, \mathbf{x}'; E) = -i(G_0^<(\mathbf{x}, \mathbf{x}'; E) + e^{\beta(E - \mu)} G_0^<(\mathbf{x}, \mathbf{x}'; E)), \quad (2.51)$$

and hence

$$G_0^<(\mathbf{x}, \mathbf{x}'; E) = i f_0(E) A(\mathbf{x}, \mathbf{x}'; E), \quad (2.52a)$$

$$G_0^>(\mathbf{x}, \mathbf{x}'; E) = -i(1 - f_0(E)) A(\mathbf{x}, \mathbf{x}'; E), \quad (2.52b)$$

where $f_0(E) = 1/(e^{\beta(E-\mu)} + 1)$ is the well-known Fermi-Dirac distribution. Eqs. (2.52a) and (2.52b) are known as the fluctuation dissipation theorem and also hold for interacting systems in equilibrium⁶.

Inserting the previous results for the retarded and advanced Green's function from Eqs. (2.41) and (2.42), Eq. (2.52a) yields an explicit expression for $G_0^<$. Making use of the Sokhotsky-Weierstrass theorem,

$$\frac{1}{E \pm i\eta} = \mathcal{P} \frac{1}{E} \mp i\pi\delta(E) \quad (2.53)$$

where \mathcal{P} denotes the Cauchy principal value integral, we arrive at

$$A(\mathbf{x}, \mathbf{x}'; E) = 2\pi \sum_n \varphi_n(\mathbf{x}) \varphi_n^*(\mathbf{x}') \delta(E - E_n), \quad (2.54)$$

and corresponding equations for $G_0^<$ and $G_0^>$. From the diagonal elements of the spectral density we find the local density of states⁷ (LDOS) [59]

$$\begin{aligned} d(\mathbf{x}, E) &= \sum_n |\varphi_n(\mathbf{x})|^2 \delta(E - E_n) = \frac{1}{2\pi} A(\mathbf{x}, \mathbf{x}; E) \\ &= \frac{i}{2\pi} (G_0^r(\mathbf{x}, \mathbf{x}; E) - (G_0^r(\mathbf{x}, \mathbf{x}; E))^*) \\ &= -\frac{1}{\pi} \Im \mathfrak{m}(G_0^r(\mathbf{x}, \mathbf{x}; E)). \end{aligned} \quad (2.55)$$

The electron density for a non-interacting system in equilibrium is then given as

$$\begin{aligned} n(\mathbf{x}) &= -i\hbar G_0^<(\mathbf{x}, \mathbf{x}; 0) = -\frac{i}{2\pi} \int_{-\infty}^{\infty} dE G_0^<(\mathbf{x}, \mathbf{x}; E) \\ &= \int_{-\infty}^{\infty} dE f_0(E) d(\mathbf{x}, E). \end{aligned} \quad (2.56)$$

Thus, in the case of an equilibrium system without interactions, the various Green's functions have a clear interpretation: The retarded Green's function contains the spectrum, i.e. the eigenfunctions and -energies, whereas the lesser Green's function describes how this spectrum is filled with particles. Even in a non-equilibrium situation, including interactions to some degree, this interpretation still holds, as is discussed below. However, the system is not described by a single Fermi-Dirac distribution anymore, and Eqs. (2.52) are replaced by a more complicated relation. The derivation of this relation for $G^<$ is the subject of the next section.

⁶ The proof, as presented here, carries over to the interacting case unchanged [64].

⁷ The local density of states, $d(\mathbf{x}, E) = \sum_n |\varphi_n(\mathbf{x})|^2 \delta(E - E_n)$ is the usual density of states, $\sum_n \delta(E - E_n)$, weighted by the spatial probability density $|\varphi_n(\mathbf{x})|^2$ of the respective electronic level.

2.4 Non-equilibrium perturbation theory

2.4.1 The need for a non-equilibrium theory

The non-equilibrium perturbation theory has been originally developed by Kadanoff and Baym [64], and from a complementary point of view by Keldysh⁸ [65], and independently by Craig [71]. Since then, it has been the subject of many reviews and books [59, 61, 63, 66, 67, 72, 73].

The basic idea of perturbation theory, equilibrium and non-equilibrium, is to express operators with a time dependence governed by the full Hamiltonian \mathcal{H} , such as $\psi_{\mathcal{H}}(\mathbf{x}, t)$, in terms of operators governed only by the non-interacting Hamiltonian \mathcal{H}_0 , such as $\psi_{\mathcal{H}_0}(\mathbf{x}, t)$. Whereas the time dependence of $\psi_{\mathcal{H}}(\mathbf{x}, t)$ is in general not known, the time dependence of $\psi_{\mathcal{H}_0}(\mathbf{x}, t)$ can be calculated, as seen in the previous paragraph.

When the time dependence of an operator is only governed by the non-interacting part \mathcal{H}_0 ,

$$A_{\mathcal{H}_0}(t) = e^{\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} A_S e^{-\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} \quad (2.57)$$

instead of the full Hamiltonian \mathcal{H} , it is said to be in the interaction picture. Formally, the Heisenberg and the interaction picture are connected by

$$A_{\mathcal{H}}(t) = u^\dagger(t, t_0) A_{\mathcal{H}_0}(t) u(t, t_0), \quad (2.58)$$

where

$$u(t, t_0) = e^{\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} U(t, t_0) \quad (2.59)$$

is the time-evolution operator in the interaction picture. This operator obeys the equation of motion

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} u(t, t_0) &= -e^{\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} \mathcal{H}_0 U(t, t_0) + e^{\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} \mathcal{H}(t) U(t, t_0) \\ &= -e^{\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} \mathcal{H}_0 e^{-\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} u(t, t_0) \\ &\quad + e^{\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} (\mathcal{H}_0 + \mathcal{H}'(t)) e^{-\frac{i}{\hbar} \mathcal{H}_0(t-t_0)} u(t, t_0) \\ &= \mathcal{H}'_{\mathcal{H}_0}(t) u(t, t_0), \end{aligned} \quad (2.60)$$

where we made use of the Schrödinger equation for the time evolution operator $U(t, t_0)$, Eq. (2.4). As for $U(t, t_0)$, we can formally integrate this equation as

$$u(t, t_0) = \mathcal{T} e^{-\frac{i}{\hbar} \int_{t_0}^t dt' \mathcal{H}'_{\mathcal{H}_0}(t')} \quad \text{for } t > t_0. \quad (2.61)$$

Again, \mathcal{T} denotes the time-ordering operator.

Note that the perturbation Hamiltonian in the interaction picture, $\mathcal{H}'_{\mathcal{H}_0}(t)$, itself only contains operators in the interaction picture, i.e. with a time-dependence governed by

⁸ Often, the non-equilibrium Green's function theory is thus referred to as Keldysh Green's function theory.

\mathcal{H}_0 . In principle, the complicated time dependence of the Heisenberg operators has thus been reduced to the simple time dependence of operators in the interaction picture—at the cost of a time-ordering operator and an infinite power series of operators, due to the exponential. In practice, this series is often only considered up to some order, in the form of a perturbation theory.

Before pursuing further into this direction, we first collect some useful identities for $u(t, t_0)$. Taking the Hermitian conjugate of Eq.(2.61), we obtain

$$u^\dagger(t, t_0) = \tilde{\mathcal{T}} e^{-\frac{i}{\hbar} \int_{t_0}^t dt' \mathcal{H}_{\mathcal{H}_0}(t')} \quad \text{for } t > t_0, \quad (2.62)$$

where $\tilde{\mathcal{T}}$ denotes the anti-time ordering⁹. Whereas $u(t, t_0)$ corresponds to a movement forward in time from t_0 to t within the interaction picture, $u^\dagger(t, t_0)$ takes the time back from t to t_0 . From these, we can define the time-evolution operator in the interaction picture for arbitrary times $t_1 > t_2 > t_0$ as $u(t_1, t_2) = u(t_1, t_0)u^\dagger(t_2, t_0)$ and find:

$$u(t_1, t_2)u^\dagger(t_1, t_2) = u^\dagger(t_1, t_2)u(t_1, t_2) = \mathbb{1} \quad \text{for } t_1 > t_2, \quad (2.63a)$$

$$u(t_1, t_2)u(t_2, t_3) = u(t_1, t_3) \quad \text{for } t_1 > t_2 > t_3, \quad (2.63b)$$

$$u^\dagger(t_1, t_2)u(t_1, t_3) = u(t_2, t_3) \quad \text{for } t_1 > t_2 > t_3. \quad (2.63c)$$

Up to now, we have only collected ideas that are also used in the well-established ground state Green's function formalism, such as presented in [58]. Why, respectively when is a non-equilibrium theory needed?

In order to answer this question, we follow an argument by Craig [71]. Suppose $|\psi_n(t_0)\rangle$ denotes a non-interacting many-body quantum state for $t_0 \rightarrow -\infty$. The matrix element of an operator with respect to this state is given as

$$\begin{aligned} \langle \psi_n(t_0) | A_{\mathcal{H}}(t) | \psi_n(t_0) \rangle &= \langle \psi_n(t_0) | u^\dagger(t, t_0) A_{\mathcal{H}_0}(t) u(t, t_0) | \psi_n(t_0) \rangle \\ &= \langle \psi_n(t_0) | u^\dagger(t_\alpha, t_0) u(t_\alpha, t) A_{\mathcal{H}_0}(t) u(t, t_0) | \psi_n(t_0) \rangle \\ &= \sum_m \langle \psi_n(t_0) | u^\dagger(t_\alpha, t_0) | \psi_m(t_0) \rangle \times \\ &\quad \langle \psi_m(t_0) | u(t_\alpha, t) A_{\mathcal{H}_0}(t) u(t, t_0) | \psi_n(t_0) \rangle, \end{aligned} \quad (2.64)$$

where $t_\alpha > t$ is some time later than any other time in the problem. Here, we made use of Eqs. (2.63) and inserted a complete set of states. If $|\psi_n(t_0)\rangle$ is a non-degenerate state, and interactions are switched on adiabatically, the time propagator in the interaction picture (that contains the interaction Hamiltonian) cannot cause transitions between states, and

$$\langle \psi_n(t_0) | u^\dagger(t_\alpha, t_0) | \psi_m(t_0) \rangle = \langle \psi_n(t_0) | u^\dagger(t_\alpha, t_0) | \psi_n(t_0) \rangle \times \delta_{nm}. \quad (2.65)$$

This is the case, when $|\psi_m(t_0)\rangle = |\Phi_0\rangle$, where $|\Phi_0\rangle$ is the non-degenerate ground state of the non-interacting system. Furthermore, from $u^\dagger(t_\alpha, t_0)u(t_\alpha, t_0) = \mathbb{1}$ and Eq. (2.65) we obtain

$$\langle \psi_n(t_0) | u^\dagger(t_\alpha, t_0) | \psi_n(t_0) \rangle = (\langle \psi_n(t_0) | u(t_\alpha, t_0) | \psi_n(t_0) \rangle)^{-1}. \quad (2.66)$$

⁹ Taking the Hermitian conjugate of a product of operators reverses the order of the operators. Thus, the Hermitian conjugate of a time-ordered product of operators is anti-time-ordered.

In the limit $t_\alpha \rightarrow \infty$ and $t_0 \rightarrow -\infty$, we then arrive at

$$\langle A_{\mathcal{H}}(t) \rangle = \frac{\langle \Phi_0 | u(\infty, t) A_{\mathcal{H}_0}(t) u(t, -\infty) | \Phi_0 \rangle}{\langle \Phi_0 | u(\infty, -\infty) | \Phi_0 \rangle} \quad (2.67)$$

which is the foundation of the ground state Green's function theory [58].

However, this is only true for a non-degenerate state and an adiabatic time dependence of the interactions. There may very well be transitions between states for finite temperatures, as excited states are commonly degenerate, and for non-equilibrium situations, where transitions between states happen by definition. In these cases, Eq. (2.65) does not hold, and it is not possible to eliminate $u^\dagger(t, t_0)$ from the equations.

In the general case, we have to take a non-interacting state in the far past, evolve it in time up to t , act on this evolved state and then bring it back into the far past again, before taking the overlap with the non-interacting state again. This is the essential content of Eq. (2.64),

$$\langle \psi_n(t_0) | A_{\mathcal{H}}(t) | \psi_n(t_0) \rangle = \langle \psi_n(t_0) | u^\dagger(t, t_0) A_{\mathcal{H}_0}(t) u(t, t_0) | \psi_n(t_0) \rangle . \quad (2.68)$$

In contrast, in the ground state formalism, the ground state evolves *uniquely* in time, and as the interaction is switched off again in the far future, we arrive again at the ground state¹⁰. In the ground state Green's function theory it therefore suffices to only consider time-evolution in one direction, whereas the non-equilibrium theory necessarily involves going forward and backward in time, as no state in the future can be uniquely identified with a state in the past. The next section will render these arguments into a rigorous theory.

2.4.2 Contour-ordered Green's function theory

We now develop a non-equilibrium Green's function theory taking into account the arguments of the preceding section, and start by considering the causal Green's function G^c that is the foundation of a perturbation theory in the ground state formalism. Without loss of generality, we assume $t > t'$ and find

$$\begin{aligned} G^c(\mathbf{x}, t, \mathbf{x}', t') &= -\frac{i}{\hbar} \langle \mathcal{T} \psi_{\mathcal{H}}(\mathbf{x}, t) \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t') \rangle \\ &= -\frac{i}{\hbar} \langle \psi_{\mathcal{H}}(\mathbf{x}, t) \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t') \rangle \\ &= -\frac{i}{\hbar} \langle u^\dagger(t, -\infty) \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, -\infty) u^\dagger(t', -\infty) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u(t', -\infty) \rangle \\ &= -\frac{i}{\hbar} \langle u^\dagger(t, -\infty) \underbrace{u^\dagger(\infty, t) u(\infty, t)}_{=1} \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, t') \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u(t', -\infty) \rangle \\ &= -\frac{i}{\hbar} \langle u^\dagger(\infty, -\infty) \times u(\infty, t) \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, t') \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u(t', -\infty) \rangle , \end{aligned} \quad (2.69)$$

¹⁰Up to some infinite phase. For details on the ground state formalism, see Ref. [58].

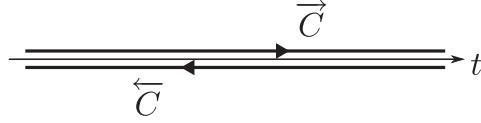


Figure 2.1 – The Keldysh contour

where we made extensive use of Eqs. (2.63). Note that all operators in the second term are time-ordered. We can therefore simplify it as

$$\begin{aligned}
 u(\infty, t) \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, t') \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u(t', -\infty) &= \\
 = \mathcal{T} \{ u(\infty, t) \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, t') \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u(t', -\infty) \} & \\
 = \mathcal{T} \{ u(\infty, -\infty) \psi_{\mathcal{H}_0}(\mathbf{x}, t) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') \} & \\
 = \mathcal{T} \{ e^{-\frac{i}{\hbar} \int_{-\infty}^{\infty} dt_1 \mathcal{H}_{\mathcal{H}_0}(t_1)} \psi_{\mathcal{H}_0}(\mathbf{x}, t) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') \}. & \quad (2.70)
 \end{aligned}$$

Here, we made use of the fact that the operators commute within the time-ordering operator, and that the time-evolution operator u always contains an even number of Fermion operators. Thus, no additional minus signs appear when interchanging the operators. Using Eq. (2.62), we finally arrive at

$$\begin{aligned}
 G^c(\mathbf{x}, t, \mathbf{x}', t') &= \\
 = -\frac{i}{\hbar} \langle \tilde{\mathcal{T}} \{ e^{-\frac{i}{\hbar} \int_{-\infty}^{\infty} dt_1 \mathcal{H}_{\mathcal{H}_0}(t_1)} \} \mathcal{T} \{ e^{-\frac{i}{\hbar} \int_{-\infty}^{\infty} dt_1 \mathcal{H}_{\mathcal{H}_0}(t_1)} \psi_{\mathcal{H}_0}(\mathbf{x}, t) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') \} \rangle. & \quad (2.71)
 \end{aligned}$$

In the ground state theory, we would only have the second, time-ordered term. However, we can *formally* bring the result (2.71) into a form equivalent to the ground state Green's function formalism, by introducing the ordering \mathcal{T}_C along a contour C , with a part \vec{C} from $-\infty$ to ∞ , and a part \overleftarrow{C} from ∞ to $-\infty$, as depicted in Fig. 2.1. The contour C is often referred to as the *Keldysh contour*.

We will in general denote a time on the contour C as τ , and the corresponding real time as t . When it is necessary to distinguish whether a real time variable t belongs to the upper or lower branch, i.e. $t \in \vec{C}$ or $t \in \overleftarrow{C}$, we will use the notation t^\rightarrow and t^\leftarrow to indicate the respective branch. Note that ordering along the contour \vec{C} corresponds to the normal time-ordering, whereas ordering along \overleftarrow{C} corresponds to anti-time ordering. Therefore we find:

$$\begin{aligned}
 G^c(\mathbf{x}, t, \mathbf{x}', t') &= \\
 = -\frac{i}{\hbar} \langle \mathcal{T}_{\vec{C}} \{ e^{-\frac{i}{\hbar} \int_{\vec{C}} dt_1 \mathcal{H}_{\mathcal{H}_0}(t_1)} \} \mathcal{T}_{\overleftarrow{C}} \{ e^{-\frac{i}{\hbar} \int_{\overleftarrow{C}} dt_1 \mathcal{H}_{\mathcal{H}_0}(t_1)} \psi_{\mathcal{H}_0}(\mathbf{x}, t^\rightarrow) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t'^\rightarrow) \} \rangle & \\
 = -\frac{i}{\hbar} \langle \mathcal{T}_C \{ e^{-\frac{i}{\hbar} \int_C d\tau_1 \mathcal{H}_{\mathcal{H}_0}(\tau_1)} \psi_{\mathcal{H}_0}(\mathbf{x}, t^\rightarrow) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t'^\rightarrow) \} \rangle & \quad \text{for } t, t' \in \vec{C}. \quad (2.72)
 \end{aligned}$$

Note that the definition of the contour-ordering operator \mathcal{T}_C also includes a minus sign for every interchange of fermionic operators.

Up to now, the contour-ordering is only a neat trick to compactify the notation: Eq. (2.72) is only a formal paraphrase of Eq. (2.71). However, the power of the contour ordering lies in the possibility to describe several Green's functions at once in a unified way. In order to demonstrate this, we next consider the lesser Green's function:

$$\begin{aligned}
G^<(\mathbf{x}, t, \mathbf{x}', t') &= \frac{i}{\hbar} \langle \psi_{\mathcal{H}}^\dagger(\mathbf{x}', t') \psi_{\mathcal{H}}(\mathbf{x}, t) \rangle \\
&= \frac{i}{\hbar} \langle u^\dagger(t', -\infty) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u(t', -\infty) u^\dagger(t, -\infty) \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, -\infty) \rangle \\
&= \frac{i}{\hbar} \langle u^\dagger(t', -\infty) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u(t', -\infty) u^\dagger(\infty, -\infty) u(\infty, -\infty) \\
&\quad u^\dagger(t, -\infty) \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, -\infty) \rangle \\
&= \frac{i}{\hbar} \langle u^\dagger(t', -\infty) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t') u^\dagger(\infty, t') u(\infty, t) \psi_{\mathcal{H}_0}(\mathbf{x}, t) u(t, -\infty) \rangle \\
&= \frac{i}{\hbar} \langle \mathcal{T}_{\overrightarrow{C}} \{ e^{-\frac{i}{\hbar} \int_{\overleftarrow{C}} dt_1 \mathcal{H}_{\mathcal{H}_0}'(t_1)} \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t'^-) \} \mathcal{T}_{\overrightarrow{C}} \{ e^{-\frac{i}{\hbar} \int_{\overleftarrow{C}} dt_1 \mathcal{H}_{\mathcal{H}_0}'(t_1)} \psi_{\mathcal{H}_0}(\mathbf{x}, t^-) \} \} \\
&= -\frac{i}{\hbar} \langle \mathcal{T}_C \{ e^{-\frac{i}{\hbar} \int_C d\tau_1 \mathcal{H}_{\mathcal{H}_0}'(\tau_1)} \psi_{\mathcal{H}_0}(\mathbf{x}, t^-) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', t'^-) \} \rangle \quad \text{for } t \in \overrightarrow{C}, t' \in \overleftarrow{C}, \quad (2.73)
\end{aligned}$$

where the minus sign in the last line is due to the interchange of the field operators. *Thus, the contour-ordered expression for $G^<$ is formally identical to G^c , only the time arguments belong to different branches of the contour.*

The fact that the contour ordering provides a unified expression for several Green's functions which only contain field operators in the interaction picture motivates the definition of the contour-ordered Green's function,

$$G(\mathbf{x}, \tau, \mathbf{x}', \tau') = -\frac{i}{\hbar} \langle \mathcal{T}_C \psi_{\mathcal{H}}(\mathbf{x}, \tau) \psi_{\mathcal{H}}^\dagger(\mathbf{x}', \tau') \rangle \quad \text{for } \tau, \tau' \in C. \quad (2.74)$$

This contour-ordered Green's function contains several of the already known Green's functions:

$$G(\mathbf{x}, \tau, \mathbf{x}', \tau') = \begin{cases} G^c(\mathbf{x}, t, \mathbf{x}', t') & \text{for } t, t' \in \overrightarrow{C}, \\ G^<(\mathbf{x}, t, \mathbf{x}', t') & \text{for } t \in \overrightarrow{C}, t' \in \overleftarrow{C}, \\ G^>(\mathbf{x}, t, \mathbf{x}', t') & \text{for } t \in \overleftarrow{C}, t' \in \overrightarrow{C}, \\ G^{ac}(\mathbf{x}, t, \mathbf{x}', t') & \text{for } t, t' \in \overleftarrow{C}. \end{cases} \quad (2.75)$$

Here, G^{ac} is the anti-causal (anti-time-ordered) Green's function.

Analogous to the derivations above, we may derive an expression for the contour-ordered Green's function only in terms of field operators in the interaction picture

$$G(\mathbf{x}, \tau, \mathbf{x}', \tau') = -\frac{i}{\hbar} \frac{\langle \mathcal{T}_C \{ e^{-\frac{i}{\hbar} \int_C d\tau_1 \mathcal{H}_{\mathcal{H}_0}'(\tau_1)} \psi_{\mathcal{H}_0}(\mathbf{x}, \tau) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', \tau') \} \rangle}{\langle \mathcal{T}_C \{ e^{-\frac{i}{\hbar} \int_C d\tau_1 \mathcal{H}_{\mathcal{H}_0}'(\tau_1)} \} \rangle} \quad \text{for } \tau, \tau' \in C. \quad (2.76)$$

Note that the denominator in this expression is simply unity. This expression allows for the formulation of a perturbation theory analogous to the ground state formalism¹¹. In particular, the Green's function can be expressed in the form of a Dyson equation [58],

$$G(\mathbf{x}, \tau, \mathbf{x}', \tau') = G_0(\mathbf{x}, \tau, \mathbf{x}', \tau') + \int_C d\tau_1 \int_C d\tau_2 \int d\mathbf{x}_1 d\mathbf{x}_2 G_0(\mathbf{x}, \tau, \mathbf{x}_1, \tau_1) \tilde{\Sigma}(\mathbf{x}_1, \tau_1, \mathbf{x}_2, \tau_2) G(\mathbf{x}_2, \tau_2, \mathbf{x}', \tau'), \quad (2.78)$$

where $\tilde{\Sigma}$ is the proper self-energy consisting only of irreducible diagrams [58]. For the non-equilibrium theory, it is useful to divide the proper self-energy $\tilde{\Sigma}$ into two parts,

$$\tilde{\Sigma}(\mathbf{x}_1, \tau_1, \mathbf{x}_2, \tau_2) = V(\mathbf{x}_1, \mathbf{x}_2, \tau_1) \delta_C(\tau_1 - \tau_2) + \Sigma(\mathbf{x}_1, \tau_1, \mathbf{x}_2, \tau_2), \quad (2.79)$$

where δ_C is the delta function on the contour C . Here, V depends only on at most one time on the contour C , whereas Σ depends explicitly on two times on C . Such a V may

¹¹Eq. (2.76) is formally equivalent to the corresponding expression of the ground state Green's function formalism, Eq. (2.67). Thus, we will not present a detailed derivation of a non-equilibrium perturbation theory here, but merely outline briefly how this perturbation arises from Eq. (2.76). All the technical details are equivalent to the ground state formalism, and the reader is referred to the extensive literature reviewing this subject [58, 62, 74].

As for the ground state, a perturbation theory is developed by expanding the exponential

$$e^{-\frac{i}{\hbar} \int_C d\tau_1 \mathcal{H}'_{\mathcal{H}_0}(\tau_1)} = 1 - \frac{i}{\hbar} \int_C d\tau_1 \mathcal{H}'_{\mathcal{H}_0}(\tau_1) + \frac{1}{2} \frac{1}{\hbar^2} \int_C \int_C d\tau_1 d\tau_2 \mathcal{H}'_{\mathcal{H}_0}(\tau_1) \mathcal{H}'_{\mathcal{H}_0}(\tau_2) + \dots, \quad (2.77)$$

as a power series in terms of the perturbing Hamiltonian $\mathcal{H}_{\mathcal{H}_0}(\tau)$, and thus also in terms of field operators $\psi_{\mathcal{H}_0}$, $\psi_{\mathcal{H}_0}^\dagger$ in the interaction picture. The contour-ordered Green's function is then given as a sum of expectation values involving a contour-ordered product of an even numbers of field operators, $\langle \mathcal{T}_C \psi_{\mathcal{H}_0}^{(\dagger)}(\mathbf{x}_1, \tau_1) \dots \psi_{\mathcal{H}_0}^{(\dagger)}(\mathbf{x}_n, \tau_n) \rangle$, and a number of contour integrations \int_C . Invoking Wick's theorem, these expectation values are then reduced to a product of expectation values involving only two field operators, $\langle \mathcal{T}_C \psi_{\mathcal{H}_0}(\mathbf{x}, \tau) \psi_{\mathcal{H}_0}^\dagger(\mathbf{x}', \tau') \rangle = G_0(\mathbf{x}, \tau, \mathbf{x}', \tau')$. Thus, the Green's function of the interacting system, G , can be expressed solely in terms of Green's functions of the non-interacting system, G_0 . The systematics of the power series expansion of $e^{-\frac{i}{\hbar} \int_C d\tau_1 \mathcal{H}'_{\mathcal{H}_0}(\tau_1)}$ together with Wick's theorem is often expressed in terms of Feynman diagrams which are a convenient way of graphically representing all the different terms arising in this expansion. Because of the formal equivalence of the contour-ordered Green's function to the ground state formalism, the Feynman rules for the contour-ordered Green's function are simply obtained by replacing real time variables with a time variable on the Keldysh contour C . Thus, all results from the ground state formalism, such as the resummation of diagrams in the form of a Dyson equation, can also be applied to the non-equilibrium formalism.

Finally, a comment on the applicability of Wick's theorem to a contour-ordered product of operators. Note that there are two versions of Wick's theorem: an operator identity that is only applicable to the ground state formalism and an identity involving only expectation values, that also holds for example in the finite-temperature formalism. For constructing a perturbation theory, it is enough to consider expectation values of contour-ordered products. In order to prove Wick's theorem it is therefore more convenient to follow the proof of Wick's theorem for finite temperatures, as given for example in chapter 24 of Ref. [58]. Indeed, this proof only requires some time-like ordering and can be applied directly to the non-equilibrium case.

arise for example from a time-dependent or time-independent potential, or within the Hartree-Fock approximation to the Coulomb-interaction [58]. In contrast to Σ , which is an expression involving two-time Green's functions, V only contains potentials or observables such as the particle density, arising from Green's functions with equal time arguments in the perturbation expansion. The Dyson equation then takes the form

$$G = G_0 + G_0 V G + G_0 \Sigma G, \quad (2.80)$$

where we compactified the notation by suppressing the arguments of the Green's function and the integrations. We will continue to do so in the following, as long as this compactified notation is unambiguous. The Dyson equation can also be written in an alternative form as

$$G = G_0 + G V G_0 + G \Sigma G_0. \quad (2.81)$$

In the ground state formalism, the distinction between V and Σ is not important. However, it will be in the case of the non-equilibrium theory, when the perturbative terms in the contour-ordered Green's function are recast into a form involving only real-time Green's functions, such as $G^<$ and G^r ,¹² as presented in the next section.

2.4.3 Real-time formulation

The contour-ordered Green's function allows for the development of a perturbation theory, but does not have any clear physical meaning. It is thus advantageous to rewrite the Dyson equations (2.80) and (2.81) in terms of the real-time Green's functions G^r , G^a , $G^<$, and $G^>$ that can be directly used, e.g. for calculating an observable. In order to do so, we follow an approach by Langreth [72, 75].

The perturbation expansion contains terms that depend on two times on the contour: These can be single contour-ordered Green's functions, or products of contour-ordered Green's functions with internal times integrated out. In order to treat these terms on equal footing, it is useful to consider a general two-time function $A(\tau, \tau')$ on the contour, $\tau, \tau' \in C$. For such function, we define the corresponding real-time functions

$$A(\tau, \tau') = \begin{cases} A^c(t, t') & \text{for } t, t' \in \vec{C}, \\ A^<(t, t') & \text{for } t \in \vec{C}, t' \in \overleftarrow{C}, \\ A^>(t, t') & \text{for } t \in \overleftarrow{C}, t' \in \vec{C}, \\ A^{ac}(t, t') & \text{for } t, t' \in \overleftarrow{C}, \end{cases} \quad (2.82)$$

$$A^r(t, t') = \theta(t - t') [A^>(t, t') - A^<(t, t')], \text{ and} \quad (2.83)$$

$$A^a(t, t') = -\theta(t' - t) [A^>(t, t') - A^<(t, t')], \quad (2.84)$$

in analogy to the definitions (2.75), (2.19a), and (2.19b). This formal definition coincides with previous Green's function identities, if $A(\tau, \tau')$ is a single contour-ordered Green's function.

¹²In essence, this distinction between V and Σ is necessary, because the definition of “lesser” or “retarded” involves two times on the contour C .

In the Dyson equation, we encounter terms of the form

$$C(\tau_1, \tau_2) = \int_C d\tau' A(\tau_1, \tau') B(\tau', \tau_2), \quad (2.85)$$

where A , B and C are two-time functions on the contour C . Specifying which branch the times τ_1 and τ_2 belong to, we can then obtain the desired real-time functions. For example, we obtain

$$\begin{aligned} C^<(t_1, t_2) &= C(t_1^\rightarrow, t_2^\leftarrow) = \int_C d\tau' A(t_1^\rightarrow, \tau') B(\tau', t_2^\leftarrow) \\ &= \int_{-\infty}^{\infty} dt' A(t_1^\rightarrow, t'^\rightarrow) B(t'^\rightarrow, t_2^\leftarrow) + \int_{\infty}^{-\infty} dt' A(t_1^\rightarrow, t'^\leftarrow) B(t'^\leftarrow, t_2^\leftarrow) \\ &= \int_{-\infty}^{\infty} dt' A^c(t_1, t') B^<(t', t_2) - A^<(t_1, t') B^{ac}(t', t_2). \end{aligned} \quad (2.86)$$

Using the identities $A^c = A^< + A^r$ and $B^{ac} = B^< - B^a$, we finally arrive at

$$C^<(t_1, t_2) = \int_{-\infty}^{\infty} dt' A^r(t_1, t') B^<(t', t_2) + A^<(t_1, t') B^a(t', t_2), \quad (2.87a)$$

an expression which only depends on real-time functions and integrations. In a similar fashion, we obtain:

$$C^>(t_1, t_2) = \int_{-\infty}^{\infty} dt' A^r(t_1, t') B^>(t', t_2) + A^>(t_1, t') B^a(t', t_2) \quad (2.87b)$$

$$C^r(t_1, t_2) = \int_{-\infty}^{\infty} dt' A^r(t_1, t') B^r(t', t_2), \quad (2.87c)$$

$$C^a(t_1, t_2) = \int_{-\infty}^{\infty} dt' A^a(t_1, t') B^a(t', t_2). \quad (2.87d)$$

These identities are commonly called *analytic continuation* or *Langreth rules*. Applying the Langreth rules repeatedly, any product of contour ordered Green's functions G can be reduced to an expression involving only $G^{<,>}$ and $G^{r,a}$. In particular, for a product of three two-time functions on the contour,

$$D = \int_C \int_C A B C, \quad (2.88)$$

we find

$$D^{<,>} = A^r B^r C^{<,>} + A^r B^{<,>} C^a + A^{<,>} B^a C^a, \quad (2.89a)$$

$$D^{r,a} = A^{r,a} B^{r,a} C^{r,a}. \quad (2.89b)$$

Here we again switched to a more compact notation, where we do not explicitly write all integration and Green's function arguments.

When applying the Langreth rules to the Dyson equations (2.80) and (2.81), it is essential to realize that the term

$$\int_C d\tau_1 G_0(\tau, \tau_1) V(\tau_1) G(\tau_1, \tau') \quad (2.90)$$

is (in terms of applying the Langreth rules) of the form $C = \int_C A B$. For the retarded Green's function, we then obtain the Dyson equations

$$G^r = G_0^r + G_0^r V G^r + G_0^r \Sigma^r G^r \quad (2.91)$$

and

$$G^r = G_0^r + G^r V G_0^r + G^r \Sigma^r G_0^r. \quad (2.92)$$

From the nature of the Langreth rules, i.e. that a retarded function can always be expressed only in terms of retarded functions, it might seem that this equation can only contain retarded functions, for any choice of the proper self-energy $\tilde{\Sigma}$. However, due to the nature of the perturbation expansion, in general there are terms containing $G^<$ or $G^>$ already in the contour-ordered expression¹³, and thus V and Σ^r may also contain $G^<$ and $G^>$.

For $G^<$, we obtain from the Langreth rules

$$G^< = G_0^< (1 + V G^a + \Sigma^a G^a) + G_0^r \Sigma^< G^a + (G_0^r V + G_0^r \Sigma^r) G^<. \quad (2.93)$$

Using Eq. (2.92) and $(1 + G^r V + G^r \Sigma^r)(1 + G_0^r V + G_0^r \Sigma^r) = 1$, we finally obtain

$$G^< = (1 + G^r V + G^r \Sigma^r) G_0^< (1 + V G^a + \Sigma^a G^a) + G^r \Sigma^< G^a. \quad (2.94)$$

Commonly, the first term of this expression is neglected: From the Dyson equation for the retarded function (2.92) and equivalently for the advanced function we can rewrite this term as $G^r \overset{\leftarrow}{G}_0^{-1} G_0^< \vec{G}_0^{-1} G^a$, where the arrows indicate on which function G_0^{-1} acts. Reversing this operator by partial integration then yields a term $\vec{G}_0^{-1} G_0^< = 0$, according to Eq. (2.38), and thus only a boundary term at times $\pm\infty$ is left. This boundary term can usually be neglected, as the initial correlations decay with time [76], except for certain cases, that we will comment on further in the next section. Hence, we obtain a compact expression for $G^<$,

$$G^< = G^r \Sigma^< G^a, \quad (2.95)$$

which is commonly referred to as the *Keldysh equation*.

¹³The perturbation expansion contains terms of the form $\mathcal{T}_C H'_{\mathcal{H}_0}(\tau_1) \dots H'_{\mathcal{H}_0}(\tau_n) \psi_{\mathcal{H}_0}(\tau) \psi_{\mathcal{H}_0}^\dagger(\tau')$. The perturbation Hamiltonian $H'_{\mathcal{H}_0}(\tau_1)$ itself contains several field operators, *at the same time* τ_1 . These cannot be ordered by T_C , but are kept as they are ordered in the expression for $H'_{\mathcal{H}_0}(\tau_1)$. Then, when applying Wick's theorem, field operators at the same time are paired, and the result is $G_0^<$ or $G_0^>$. In Feynman diagrams, such terms appear, for example, in the form of Fermion bubbles, and are already present in the ground state formalism [58].

Up to now, all expressions were for Green's functions in the time domain. In steady-state, i.e. when the properties of the (non-equilibrium) system are constant in time and the Green's functions only depend on time differences, it is more convenient to consider the real-time Green's functions in the energy domain. Note that this is not possible for the contour-ordered Green's function, as the contour C has two different branches.

Working in the energy domain is especially useful for steady state, as all expressions for real-time Green's functions involve convolutions that are reduced to simple multiplications: The Fourier transform of $A(t - t') = \int_{-\infty}^{\infty} dt_1 B(t - t_1)C(t_1 - t')$ yields $A(E) = B(E)C(E)$. This gives, for example, the Keldysh equation in the energy domain as

$$G^<(E) = G^r(E) \Sigma^<(E) G^a(E), \quad (2.96)$$

where again we suppressed the remaining spatial integrations. Our compact notation is particularly useful in this regard: All expressions can be understood to be valid both for the time and the energy domain. In the remainder of this work, we restrict ourselves to steady state and work exclusively in the energy domain.

During the course of the discussion about the Green's function theory, we have introduced the self-energy contributions V and Σ as general concepts. In the next section, we will derive explicit expressions for these objects.

2.5 Transport in tight-binding models

2.5.1 Green's functions in tight-binding

The non-equilibrium Green's function formalism of the previous chapter in principle provides a systematic prescription to perform a transport calculation for any system. However, even in the case of a non-interacting system, an exact analytic solution is usually not possible. Therefore, numerical computations are often the method of choice.

As presented above, the Green's functions obey differential or integral equations and depend on variables \mathbf{x} , \mathbf{x}' with a continuous range of values. These continuous degrees of freedom are difficult to treat on a computer, and thus a numerical solution is in general only attempted within a *discrete* basis set $\{|i\rangle\}$. In this basis set, the single-particle Hamiltonian can be written as

$$H = \sum_{ij} H_{ij} |i\rangle \langle j|, \quad (2.97)$$

where $H_{ij} = \langle i | H | j \rangle$. Very often, only few of the matrix elements H_{ij} are nonzero. Such *tight-binding* representations of the Hamiltonian are ubiquitous in quantum transport calculations and can arise from finite differences [77–79], from the finite element method [80], from atomic orbitals in empirical tight-binding [81–83] or Kohn-Sham orbitals within density functional theory [84–86]. In this work, we will employ both the finite difference approximation to the effective mass Schrödinger equation (Chapter

5 and Appendix F) and an empirical tight-binding model for graphene (Chapters 6–7 and Appendix G). In both cases, the states $|i\rangle$ form an orthonormal basis. In the following we therefore restrict the discussion to orthogonal tight-binding models.

The corresponding non-interacting many-particle Hamiltonian is then given as

$$\mathcal{H} = \sum_{ij} H_{ij} c_i^\dagger c_j \quad (2.98)$$

and is simply obtained by replacing the states $|i\rangle$ with the corresponding annihilation and creation operators [70]. Instead of the real-space Green's functions, that are given as expectation values of two field operators $\psi(\mathbf{x})$ and $\psi^\dagger(\mathbf{x}')$, we now consider the Green's functions in the representation of the discrete basis set $|i\rangle$. These are obtained by replacing $\psi(\mathbf{x}) \rightarrow c_i$ and $\psi^\dagger(\mathbf{x}') \rightarrow c_j^\dagger$. For example, the lesser Green's function in the basis set representation is then given as

$$G^<(i, t, j, t') = \frac{i}{\hbar} \langle c_j^\dagger(t') c_i(t) \rangle. \quad (2.99)$$

Note that in many cases, the basis states $|i\rangle$ are sufficiently localized at a point \mathbf{x}_i . In these cases, c_i plays the role of $\psi(\mathbf{x}_i)$ and the interpretation of the Green's functions remains unchanged. However, in any case, the Green's functions in tight-binding representation can be used to calculate any desired observable (see Appendix A).

Because of the discreteness of the basis set, the Green's functions can be interpreted as matrices, and we will use the notation $G_{ij}(t, t') = G(i, t, j, t')$ or $G_{ij}(E) = G(i, j; E)$ to emphasize this fact. All the results of the previous section also carry over to the Green's function in basis set representation, with integrations over \mathbf{x} replaced by sums over basis states i , $\int d\mathbf{x}_1 A(\mathbf{x}, \mathbf{x}_1) B(\mathbf{x}_1, \mathbf{x}') \rightarrow \sum_k A_{ik} B_{kj}$. Therefore, in tight-binding representation, the results of the previous section can be directly interpreted as matrix equations. This is another advantage of the compact notation employed there.

2.5.2 Transport equations

We consider a non-interacting system consisting of a *scattering region* (also called central region) and N_L *leads*, as schematically depicted in Fig. 2.2. The leads and the scattering region are assumed to be disconnected in the far past, and the hopping is switched on adiabatically at later times [61]. The Hamiltonian of the system is then given as

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}'. \quad (2.100)$$

Here, \mathcal{H}_0 is the Hamiltonian of the disconnected subsystems,

$$\mathcal{H}_0 = \sum_{l=1}^{N_L} \mathcal{H}_l + \mathcal{H}_S, \quad (2.101)$$

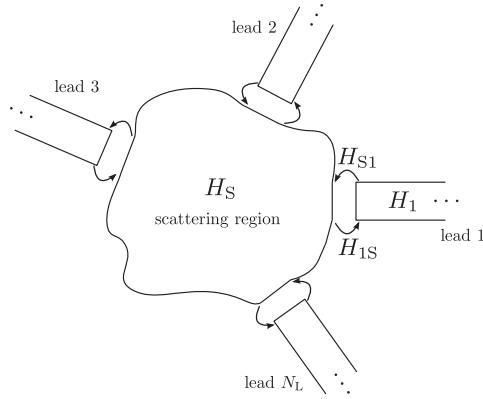


Figure 2.2 – In a transport calculation, the system under consideration consists in general of a scattering region attached to a number of leads.

where \mathcal{H}_l is the Hamiltonian of lead l and \mathcal{H}_S of the scattering region, and \mathcal{H}' gives the coupling of the leads to the scattering region,

$$\mathcal{H}' = \sum_{l=1}^{N_L} \mathcal{H}_{lS} + \mathcal{H}_{Sl} \quad (2.102)$$

where \mathcal{H}_{lS} and \mathcal{H}_{Sl} describe the hopping from the scattering region to lead l and vice versa. Since \mathcal{H} is a Hermitian Hamiltonian, we have $\mathcal{H}_{lS} = \mathcal{H}_{Sl}^\dagger$. The Hamiltonians of the isolated parts and of the hopping are given as

$$\mathcal{H}_S = \sum_{ij} H_{S,ij} c_{S,i}^\dagger c_{S,j}, \quad (2.103)$$

$$\mathcal{H}_l = \sum_{ij} H_{l,ij} c_{l,i}^\dagger c_{l,j}, \quad (2.104)$$

$$\mathcal{H}_{lS} = \sum_{ij} H_{lS,ij} c_{l,i}^\dagger c_{S,j}. \quad (2.105)$$

Here, the creation and annihilation operators are marked with a subscript, depending on whether they act in the scattering region ($c_{S,i}$, $c_{S,i}^\dagger$) or in lead l ($c_{l,i}$, $c_{l,i}^\dagger$). Note that the Hamiltonian is marked with a single subscript S or l only, if it contains operators within a single subsystem, and with two subscripts otherwise. We will employ the same notation for the Green's function later.

In the far past, the isolated subsystems are assumed to be in thermal equilibrium with electrochemical potential μ_l in lead l and μ_S in the scattering region¹⁴. The

¹⁴The electrochemical potential of the leads μ_l is determined by the number of particles in lead l and

electrochemical potentials μ_l may be different for every lead, and this difference will be the driving force for non-equilibrium. Such a difference may be for example due to different electron densities, temperatures or electrical potentials in the leads. Here, we consider in particular the case of an applied bias, such that $\mu_l = \mu_0 + qV_l$, where V_l is the electrical potential in lead l , q the charge of the particles, and μ_0 the equilibrium chemical potential. In the following, we will specialize to electronic transport, where $q = -e$ is the charge of an electron.

We now apply the results of the non-equilibrium Green's function formalism of the previous section for the perturbation Hamiltonian \mathcal{H}' and obtain the Dyson equation for the contour-ordered Green's function,

$$G = G_0 + G_0 V G = G_0 + G V G_0 \quad (2.106)$$

where $V = \sum_l H_{lS} + H_{Sl}$. Note that this equation is *exact* to all orders in the perturbation, as \mathcal{H}' is simply a (off-diagonal) potential, and the proper self-energy given by a single term V [60].

In particular, we are interested in the Green's function within the scattering region, G_S . Since the leads and the scattering region are decoupled in the far past, there are no correlations between them, and as a consequence $G_{0,Sl} = G_{0,lS} = 0$. Using this fact, we find

$$G_S = G_{0,S} + G_{0,S} \sum_l H_{Sl} G_{lS} \quad (2.107)$$

and

$$G_{lS} = G_{0,l} H_{lS} G_S \quad (2.108a)$$

$$G_{Sl} = G_S H_{Sl} G_{0,l} \quad (2.108b)$$

Inserting Eq. (2.108a) into (2.107), we obtain

$$G_S = G_{0,S} + G_{0,S} \Sigma G_S , \quad (2.109)$$

where $\Sigma = \sum_l H_{Sl} G_{0,l} H_{lS}$ is the self-energy due to the leads.

We can now apply the analytic continuation rules of the previous section to obtain the real-time Green's functions. For the retarded Green's function, Eq. (2.91), we find $G_S^r = G_{0,S}^r + G_{0,S}^r \Sigma^r G_S^r$. The analytic continuation of the contour-ordered self-energy Σ is particularly simple in this case, as it consists only of a single Green's function, and we obtain $\Sigma^r = \sum_l \Sigma_l^r = \sum_l H_{Sl} G_{0,l}^r H_{lS}$, where Σ_l^r is the retarded self-energy due to a single lead l . In a tight-binding model, the hopping matrices H_{lS} only couple states at the boundary between the scattering region and the leads. Thus, Σ^r only depends on

the electrical potential arising from some applied voltage. In contrast, the electrochemical potential of the scattering region μ_S is somewhat ill-defined, as there is no connection to a reservoir with a well-defined electrical potential. Fortunately, the electrochemical potential of the scattering region does not enter the final result.

the values of $G_{0,l}^r$ on the surface of lead l , decoupled from the scattering region. We denote the values of $G_{0,l}^r$ on the surface as the *surface Green's function* g_l^r , such that

$$\Sigma^r = \sum_l H_{Sl} g_l^r H_{lS}, \quad (2.110)$$

Using furthermore $G_{0,S} = (E - H_S + i\eta)^{-1}$, Eq. (2.43), we arrive at the final expression for the retarded Green's function,

$$G_S^r = (E - H_S - \Sigma^r + i\eta)^{-1}. \quad (2.111)$$

Since the scattering region is finite, so are the matrices H_S and Σ^r , and the problem of calculating the retarded Green's function has been reduced to the problem of inverting a finite matrix. The infinitesimal $i\eta$ can be neglected in a numerical calculation, if the self-energy of the leads has a nonzero imaginary part, $\Im(\Sigma^r) \neq 0$. This is the case, if there are propagating states in the leads at the energy E .

The lesser Green's function of the scattering region, Eq. (2.95), is given as

$$G_S^< = G_S^r \Sigma^< G_S^a, \quad (2.112)$$

where $\Sigma^< = \sum_l H_{Sl} g_l^< H_{lS}$. Since the leads were assumed to be in equilibrium in the far past, we can employ Eq. (2.52a) to calculate $\Sigma^<$:

$$\begin{aligned} \Sigma^< &= \sum_l H_{Sl} g_l^< H_{lS} \\ &= \sum_l H_{Sl} (-f_0(E, \mu_l)) (g_l^r - g_l^a) H_{lS} \\ &= \sum_l i f_0(E, \mu_l) \Gamma_l, \end{aligned} \quad (2.113)$$

where we defined $\Gamma_l = i(\Sigma_l^r - \Sigma_l^a) = i(\Sigma_l^r - (\Sigma_l^r)^\dagger)$.

These Green's functions can now be used to calculate the current through the system. For this, we define the current in lead l as $I_l = \frac{d}{dt} Q_l$, where Q_l is the charge in lead l . With this definition, a positive current $I_l > 0$ corresponds to an increase of positive charge, and thus to electrons flowing out of the lead. With $Q_l = -e\langle N_l \rangle$, where $N_l = \sum_i c_{l,i}^\dagger c_{l,i}$ is the number operator in lead l , we obtain through direct calculation

$$\begin{aligned} I_l &= -e \langle \frac{d}{dt} N_l \rangle = \frac{ie}{\hbar} \langle [N_l, \mathcal{H}] \rangle \\ &= \frac{ie}{\hbar} \sum_{ij} H_{lS,ij} \langle c_{l,i}^\dagger c_{S,j} \rangle - H_{Sl,ji} \langle c_{S,j}^\dagger c_{l,i} \rangle \\ &= e \sum_{ij} H_{lS,ij} G_{Sl,ji}^< - H_{Sl,ji} G_{lS,ij}^< \\ &= e \text{Tr} (H_{lS} G_{Sl}^< - H_{Sl} G_{lS}^<) . \end{aligned} \quad (2.114)$$

Here, we used the definition of the lesser Green's function, Eq. (2.19c), and the definition of the trace.

Note that the expressions are up to now still in the time-domain. In steady-state¹⁵, the current is constant in time, $I_l(t) = I_l(0) = I_l$. Hence, it is possible to work in the energy domain and write

$$I_l = I_l(0) = \frac{e}{h} \int dE \operatorname{Tr}(H_{lS} G_{Sl}^<(E) - H_{Sl} G_{lS}^<(E)) . \quad (2.115)$$

For brevity, in the following we will again omit the dependence of the Green's functions on energy E . Inserting the analytic continuation of Eq. (2.108) we obtain

$$\begin{aligned} I_l &= \frac{e}{h} \int dE \operatorname{Tr}(H_{lS} G_S^r H_{Sl} G_{0,l}^< + H_{lS} G_S^< H_{Sl} G_{0,l}^a \\ &\quad - H_{Sl} G_{0,l}^r H_{lS} G_S^< - H_{Sl} G_{0,l}^< H_{lS} G_S^a) \\ &= \frac{e}{h} \int dE \operatorname{Tr}((\Sigma_l^a - \Sigma_l^r) G_S^< + \Sigma_l^< (G_S^r - G_S^a)) \\ &= \frac{e}{h} \int dE \operatorname{Tr}(i\Gamma_l (G_S^< + f_0(E, \mu_l) (G_S^r - G_S^a))) , \end{aligned} \quad (2.116)$$

where we made use of the cyclic invariance of the trace and Eq. (2.113). The expression for the current (2.116) was first derived by Meir and Wingreen [87], and holds also for a system with interactions in the scattering region. For the case of non-interacting particles—or more precisely for systems where $\Sigma^<$ only has contributions from the leads as in Eq. (2.113)—we can use Eqs. (2.112) and (2.23) to further simplify this expression:

$$\begin{aligned} I_l &= \frac{e}{h} \int dE \operatorname{Tr}(i\Gamma_l \sum_{l'} G_S^r i f_0(E, \mu_{l'}) \Gamma_{l'} G_S^a + i f_0(E, \mu_l) \Gamma_l G_S^r (\underbrace{\Sigma^> - \Sigma^<}_{=-i \sum_{l'} \Gamma_{l'}}) G_S^a) \\ &= \frac{e}{h} \int dE \sum_{l'} \operatorname{Tr}(\Gamma_l G_S^r \Gamma_{l'} G_S^a) (f_0(E, \mu_l) - f_0(E, \mu_{l'})) \\ &= \frac{e}{h} \int dE \sum_{l'} T_{ll'}(E) (f_0(E, \mu_l) - f_0(E, \mu_{l'})) , \end{aligned} \quad (2.117)$$

where we defined $T_{ll'}(E) = \operatorname{Tr}(\Gamma_l G_S^r \Gamma_{l'} G_S^a)$. In the next section, we will give a physical meaning to this object.

¹⁵There is a seeming contradiction between the assumption of steady state, i.e. time-independent observables, and the definition of current as $I_l = \frac{d}{dt} Q_l$. Per definition, $\frac{d}{dt} Q_l = 0$ in steady-state, and thus it seems as there could not be any current under steady-state conditions. However, this is a problem of the order of taking limits: The leads are assumed to be infinitely large. Therefore, we have to take both the limit of approaching steady state ($\lim_{t_0 \rightarrow -\infty}$) and infinitely large leads, $\lim_{Q_l \rightarrow \infty}$. Physically, this means that the electrochemical potential of the leads is not changed by a current flow.

When the difference in the electrochemical potentials, i.e. the applied voltage, is small such that $\mu_l \approx \mu_{l'}$, it makes sense to define a Fermi energy $E_F \approx \mu_l$ for the whole system. The current (2.117) depends on the difference between the electrochemical potentials μ_l and $\mu_{l'}$ through the difference of the Fermi functions. Thus, only a small energy window between and around μ_l and $\mu_{l'}$ contributes to the energy integral. Provided that $T_{ll'}(E)$ does not vary significantly in this energy window, we can approximate $T_{ll'}(E) \approx T_{ll'}(E_F)$ and take it out of the energy integral. For small bias voltages and low temperatures, i.e. in the *linear response regime* [59, 60], we then find

$$\begin{aligned} I_l &\approx \frac{e}{h} \sum_{l'} T_{ll'}(E_F) \int dE (f_0(E, \mu_l) - f_0(E, \mu_{l'})) \\ &= \frac{e}{h} \sum_{l'} T_{ll'}(E_F) (\mu_l - \mu_{l'}) = \frac{e}{h} \sum_{l'} T_{ll'}(E_F) (-eV_l + eV_{l'}) \\ &= \sum_{l'} G_{ll'}(V_{l'} - V_l), \end{aligned} \quad (2.118)$$

where we defined the *conductance*

$$G_{ll'} = \frac{e^2}{h} T_{ll'}(E_F) = \frac{e^2}{\hbar} \text{Tr}(\Gamma_l G_S^r \Gamma_{l'} G_S^a) \quad (2.119)$$

between two leads l and l' .

The equations for G_S^r and $G_S^<$, (2.107) and (2.112) respectively, are the central results for calculating transport properties in a tight-binding model: The currents I_l can be calculated from G_S^r , while any local observable may be obtained from $G_S^<$ (see Appendix A). The non-equilibrium formalism thus provides a framework that can be applied to both interacting and non-interacting systems, with equations that are very suitable for a computer implementation, as they only involve matrix multiplications and inversions. For this reason the NEGF formalism has become rather popular for numerical computations, even for non-interacting systems. As a matter of fact, the non-equilibrium formalism can even be applied as a black box to obtain physical results. However, in such an approach it is difficult to develop a physical intuition of the underlying processes.

Unfortunately, the physical meaning of the Green's functions in the non-equilibrium formalism is clouded by the rather involved derivation of the basic equations. In fact, for a fully interacting system an intuitive picture may not be possible at all. This is however not true for non-interacting systems: In the next section we will give a physical meaning to the Green's functions by showing the equivalence of the NEGF and the scattering formalism for non-interacting systems, and further discuss to what extent this interpretation carries over to interacting systems.

2.5.3 Connection with the scattering formalism

A very intuitive approach to quantum transport in non-interacting systems has been pioneered by Landauer and Büttiker [88, 89]. As already observed in the case of

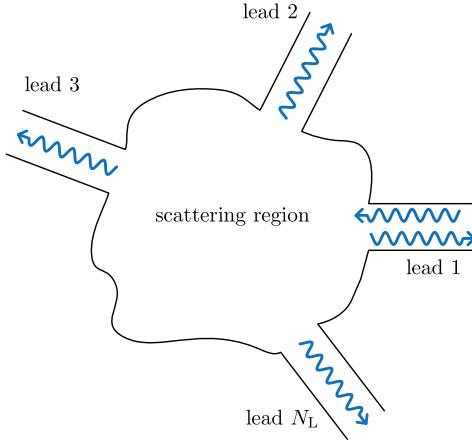


Figure 2.3 – An incoming wave from a certain lead is scattered inside the scattering region and may be either reflected back, or transmitted into another lead.

Green's functions, a non-interacting system can be characterized by the eigenstates of the single-particle Hamiltonian H . In the leads, these eigenstates are longitudinal plane waves with a transverse wave function with quantum index n (also called *channel* n ; here n includes both the orbital and spin degrees of freedom). An incoming wave is scattered within the scattering region, and can be either reflected back into the same lead or transmitted into a different lead (see Fig. 2.3). The quantum mechanical wave function that describes the scattering of an incoming wave in channel n from lead l is the *scattering wave function* $\Psi_n^{(l)}$. The central result of the Landauer-Büttiker formalism, also called scattering formalism, relates the conductance of a system to the properties of the scattering wave functions: the linear conductance between two leads is given by the sum of the transmission probabilities between the different channels in lead l and l' ,

$$G_{ll'} = \frac{e^2}{h} \sum_{mn} |t_{ll',nm}|^2 \quad (2.120)$$

where $t_{ll',nm}$ is the transmission amplitude from channel m in lead l' to channel n in lead l at the Fermi energy E_F . A detailed exhibition of the Landauer-Büttiker formalism in a tight-binding model is given in Appendix C.

Whereas the transport properties in the Landauer-Büttiker formalism are uniquely defined by the scattering wave function, they are given in terms of Green's functions in the non-equilibrium Green's function formalism. In order to connect these complementary approaches, we need a relation between the scattering wave function and the

Green's function. Such a relation is commonly called *Fisher-Lee relation*¹⁶. As shown in Appendix C for an arbitrary tight-binding model, Eq. (C.30), the scattering wave function at some (lattice) point \mathbf{x}_i in the scattering region is given as

$$\Psi_n^{(l)}(\mathbf{x}_i) \propto G_{\mathbf{x}_i, 0_l}^r \Gamma_l \phi_{n,>}^{(l)}, \quad (2.121)$$

where $\phi_{n,>}^{(l)}$ is the transverse wave function of an incoming channel n in lead l , and 0_l denotes the surface of lead l . This is the central result for showing the equivalence of the Green's function and the scattering formalism.

As shown in detail in Appendix C, we obtain

$$T_{ll'}(E_F) = \text{Tr}(\Gamma_l G_S^r \Gamma_{l'} G_S^a) = \sum_{mn} |t_{ll',nm}|^2, \quad (2.122)$$

proving the equivalence of the scattering and the non-equilibrium Green's function formalism. Thus, $T_{ll'}(E_F)$ can be interpreted as the total transmission probability of all channels in lead l' into all channels in lead l . In addition, we obtain an expression for $G_S^<$ as (see Eq. (C.38))

$$G_{S,\mathbf{x}_i \mathbf{x}_j}^<(E) \propto \sum_l f_0(E, \mu_l) \sum_n i \Psi_n^{(l)}(\mathbf{x}_i) (\Psi_n^{(l)}(\mathbf{x}_j))^* D_n^{1d,(l)}(E), \quad (2.123)$$

where $D_n^{1d,(l)}(E)$ is the one-dimensional¹⁷ density of states for a channel n in lead l . Thus, the lesser Green's function describes which scattering states are occupied at energy E . Whether a scattering wave function $\Psi_n^{(l)}$ is occupied or not is determined solely by the electrochemical potential μ_l of the respective leads.

Eq. (2.123) thus leads to a rather intuitive physical picture, depicted for the case of two leads, left (L) and right (R), in Fig. 2.4: The electronic states of the system are given by the scattering wave functions from left and right lead. For energies smaller than $\mu_R < \mu_L$, both scattering wave functions emanating from the left lead (right-moving wave functions) and from the right lead (left-moving) are occupied¹⁸. In the energy window $\mu_R < E < \mu_L$, only right-moving states are occupied, and there are no occupied states for energies above μ_L . This is a generalization of the interpretation of the lesser Green's function in Section 2.3, where we assumed a constant electrochemical potential throughout the whole system.

¹⁶It was Fisher and Lee that first published a relation between the retarded Green's function and the transmission amplitudes [90].

¹⁷A channel n has an energy dispersion $E_{n,k}$ depending on a one-dimensional real wave vector k (see Section 3.1). Thus, a single channel behaves like a one-dimensional system.

¹⁸For the sake of simplicity, we assume in our interpretation very low temperatures, $T \approx 0$, such that the Fermi function can be replaced by a step function $f_0(E, \mu) = \theta(\mu - E)$. For finite temperatures, there is some energy region around μ_R and μ_L where the smearing of the Fermi function also allows some additional occupancy.

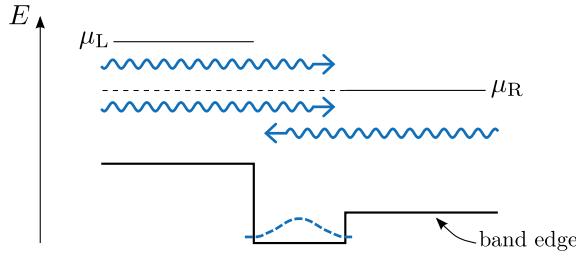


Figure 2.4 – Schematic of a two-terminal system in non-equilibrium: scattering states emanating from the left (right) lead are occupied up to the electrochemical potential μ_L (μ_R), as depicted by wavy lines. In principle, there may also be localized states, as indicated by the dashed line.

From this picture, we can also understand the appearance of the factor $f_0(E, \mu_L) - f_0(E, \mu_R)$ in the current equation (2.117): For $E < \mu_R$, where $f_0(E, \mu_L) - f_0(E, \mu_R) = 0$ the currents carried by the left- and right-moving states cancel exactly. In contrast, for $\mu_R < E < \mu_L$ there are only right-moving states leading to a finite current from left to right.

In principle, there could also be localized states in the scattering region, that are not connected to any scattering state emanating from the leads, as indicated in Fig. 2.4. For example, such a state could lie below the band edges of the leads, and since there are no transitions between states of different energy in a non-interacting system, this localized state cannot be connected to any scattering state. These states are thus *not* included in the simplified form of the Keldysh equation (2.95). Instead, they are contained in the first term of Eq. (2.94) that is usually neglected [76]. Note that the first term of Eq. (2.94) involves the lesser Green's function $G_{0,S}^<$ of the disconnected scattering region, and thus μ_S . Essentially, this means that the electrochemical potential μ_S in the far past determines whether a localized state is occupied or not. This comes at no surprise, since it cannot be filled from the states of the leads that are at a different energy. Since μ_S is somewhat ill-defined as mentioned above, additional physical arguments are required to determine the occupancy of localized states. Fortunately, localized states do not contribute to transport in non-interacting systems, as they do not couple to the scattering states, and are thus of minor importance in this work. In interacting systems however, they may influence transport through the interaction term, and additional care must be taken to account for these states [85, 91].

Transport in a non-interacting system can thus be understood in a rather intuitive picture. To which extent does this interpretation carry over to the interacting case? Eqs. (2.122) and (2.123) are of central importance in showing the equivalence of the scattering and the non-equilibrium Green's function formalism. The main assumption behind the derivation of these equations is that the lesser self-energy $\Sigma^<$ only contains contributions from the leads. This is the case in non-interacting systems, but also for

the example of Coulomb interaction within Hartree-Fock approximation and generally for any interaction in mean-field approximation (including density functional theory¹⁹). In all these cases, transport can be still understood in terms of scattering wave functions in an effective potential. Only if the interaction is treated beyond mean-field, leading to a contribution to $\Sigma^<$, this interpretation breaks down, as then (virtual) transitions between states of different energies become important.

Since the scattering formalism and the NEGF formalism for non-interacting systems are equivalent, often concepts of both approaches are used in a calculation at the same time. For example, since calculating the scattering wave function directly is in general numerically unstable [92], numerical calculations in the Landauer-Büttiker formalism usually first compute G_S^r from Eq. (2.111), and then the transmission amplitudes $t_{ll',nm}$ through the Fisher-Lee relation. Note that the $t_{ll',nm}$ contain more information than just the total transmission $T_{ll'}(E)$. Because of that, and since in addition calculating the total transmission probability $T_{ll'}(E)$ from $t_{ll',nm}$ is computationally less expensive than computing $T_{ll'}(E)$ from the Green's functions, we also employ this approach in the numerical work. The main conceptual difference between the Landauer-Büttiker formalism and the non-equilibrium Green's function formalism is the calculation of $G_S^<$, and in this sense Eq. (2.112) contains the “non-equilibrium”. As seen above, this quantity could also be calculated from the scattering wave functions, but this is usually not done. It should be emphasized that of course also the Landauer-Büttiker formalism is a non-equilibrium theory—there would be no current and hence no conductance in equilibrium. A non-equilibrium situation, i.e. different electrochemical potentials in the leads, enters explicitly in the derivation of the Landauer-Büttiker formula.

The scattering formalism has the advantage of providing a more intuitive picture of transport. In contrast, the Green's functions in the non-equilibrium Green's function formalism are very convenient to compute numerically, and the formalism can be easily extended to include interactions. Because of this the NEGF formalism has become rather popular for numerical computations.

2.6 Challenges for a numerical implementation

The transport formalism presented in the previous section is very general, as it is applicable to any tight-binding model. Central results were Eq. (2.111), $G_S^r = (E - H_S - \Sigma^r)^{-1}$, and Eq. (2.112), $G_S^< = G_S^r \Sigma^< G_S^a$. However, directly implementing these equations in a computer program still faces some difficulties. In this section, we will identify the remaining steps necessary to build a generic transport code for arbitrary

¹⁹In the case of density functional theory, the physical interpretation is complicated by the fact that the scattering wave functions are then solutions to the Kohn-Sham equations. However, the solutions of the Kohn-Sham equations do not have any physical meaning by themselves and neither do the corresponding scattering wave functions. But then, transport theory in density functional theory must assume anyway that the wave functions obtained like that are close to physical wave functions [84].

tight-binding models. In the following two chapters we will then present novel solutions to these problems.

The first step is obvious from Eq. (2.111): we need to compute Σ^r . In Chapter 3 we thus derive an expression for the self-energy Σ_l^r of a lead l , valid for arbitrary tight-binding models, paying close attention to the numerical stability of this method.

The second step is the calculation of G_S^r . At first glance, this does not appear problematic, as G_S^r can be in principle computed by direct inversion as given in Eq. (2.111). This inversion involves a matrix of size N_{grid} , where N_{grid} is the number of degrees of freedom in the tight-binding model, i.e. the number of lattice sites or orbitals in the scattering region. However, physical simulations involve in general large-scale systems, and a direct inversion is then computationally intractable.

Then again, it is often not necessary to calculate the full retarded Green's function G_S^r : For example, to obtain the conductance, it is enough to compute the Green's function between the leads, as can be seen from Eqs. (2.117) and (C.31) and the fact that the self energies from the leads Σ_l^r only couple to the surface of the scattering region. This and the fact that, in a tight-binding model, H_S is a sparse matrix, is the foundation of several quantum transport algorithms that avoid the direct inversion of Eq. (2.111). These transport algorithms however have the disadvantage of relying on some assumptions or geometrical restrictions on the underlying system, and are thus not applicable to arbitrary tight-binding models. In Chapter 4 we review several well-established transport algorithms and develop a preconditioning step that allows to apply these algorithms to an arbitrary tight-binding model. A detailed description of the quantum transport algorithm employed in this work, the *recursive Green's function* technique, is given in Appendix B.

Chapter 3

Lead Green's functions

3.1 Introduction

Calculating transport properties in the non-equilibrium Green's function formalism requires the evaluation of the lead self-energies $\Sigma_l^r = H_{Sl} g_l^r H_{lS}$, as seen in the previous chapter. Finding an expression for the surface Green's function of the lead g_l^r , and thus for Σ_l^r , is the intent of this chapter. Below, we will drop the subscript l since we are dealing here with a single lead only, and simply denote $g_l^r = g^r$. In addition, for the sake of brevity, we will often refer to the surface Green's function of the lead simply as the *lead Green's function*.

In general, a lead is a semi-infinite structure attached to the scattering region. In particular, we consider leads that in addition exhibit some periodicity, so that asymptotic scattering states can be defined. In tight-binding representation, such a lead is a semi-infinite repetition of unit cells, governed by the unit cell Hamiltonian H_0 , with neighboring unit cells connected by hopping Hamiltonians H_1 and $H_{-1} = H_1^\dagger$, as depicted in Fig. 3.1. If there are $N_{\text{u.c.}}$ degrees of freedom within a unit cell, these Hamiltonians can be represented by $N_{\text{u.c.}} \times N_{\text{u.c.}}$ matrices.

For transport, we have to calculate the surface Green's function of the lead detached from the scattering region. The Hamiltonian of such an isolated, semi-infinite lead is given by

$$H_{\text{semi-infinite}} = \begin{pmatrix} H_0 & H_1 & 0 & \dots \\ H_{-1} & H_0 & H_1 & \\ 0 & H_{-1} & H_0 & \\ \vdots & & & \ddots \end{pmatrix}, \quad (3.1)$$

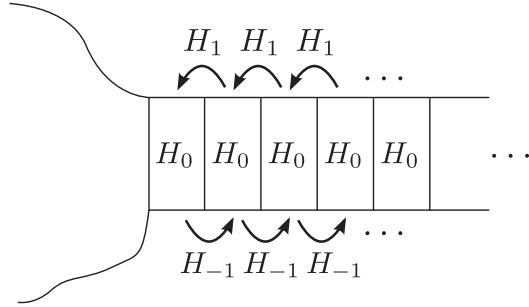


Figure 3.1 – Schematic view of a lead: A semi-infinite series of unit cells.

and we can write this Hamiltonian in a form reminiscent of a one-dimensional wire,

$$H_{\text{semi-infinite}} = \sum_{j=0}^{\infty} H_0 |j\rangle \langle j| + H_1 |j\rangle \langle j+1| + H_{-1} |j\rangle \langle j-1| , \quad (3.2)$$

where j is the unit cell index. The surface Green's function g^r is defined only within the first unit cell of the semi-infinite wire, $g^r = \langle 0 | G_{\text{semi-infinite}} | 0 \rangle$, where $G_{\text{semi-infinite}}$ is the full Green's function of the isolated lead. Therefore, the quantity that we desire to calculate is a well-defined, finite matrix and its size is given by the size of the unit cell Hamiltonian H_0 .

A problem related to that of the semi-infinite lead is the infinite wire that extends to infinity in both directions. Its Hamiltonian is given by

$$H_{\text{infinite}} = \begin{pmatrix} \dots & & & & & \\ & H_0 & H_1 & & & \\ & H_{-1} & H_0 & H_1 & & \\ & & H_{-1} & H_0 & H_1 & \\ & & & H_{-1} & H_0 & \\ & & & & & \ddots \end{pmatrix} , \quad (3.3)$$

or in more compact form

$$H_{\text{infinite}} = \sum_{j=-\infty}^{\infty} H_0 |j\rangle \langle j| + H_1 |j\rangle \langle j+1| + H_{-1} |j\rangle \langle j-1| . \quad (3.4)$$

The Hamiltonian H_{infinite} is periodic, and thus its eigenstates can be written in Bloch form [93]:

$$\varphi_{n,k}(j) = \phi_{n,k} e^{ikj} , \quad (3.5)$$

where j again denotes the unit cell index, $k \in [-\pi, \pi]$, and $\phi_{n,k}$ is a vector representing the “transverse” part of the wave function within a unit cell. The wave function $\varphi_{n,k}$

is a solution of the Schrödinger equation

$$(E_{n,k} - H_{\text{infinite}}) \varphi_{n,k} = 0 \quad (3.6)$$

if the transverse wave function $\phi_{n,k}$ obeys

$$(E_{n,k} - H_0 - H_1 e^{ik} - H_{-1} e^{-ik}) \phi_{n,k} = 0. \quad (3.7)$$

The effective Hamiltonian $H(k) = H_0 + H_1 e^{ik} + H_{-1} e^{-ik}$ is Hermitian and thus, for a given k , the wave functions $\{\phi_{n,k}\}$ form a complete orthonormal set.

The infinite wire is related to the problem of the semi-infinite lead in several ways: First, away from the contact to the scattering region, the *asymptotic* eigenstates of the semi-infinite lead are given by the eigenstates $\varphi_{n,k}$ of the infinite wire. Furthermore, the Green's functions of the semi-infinite lead and the infinite wire are connected via Dyson's equation, as discussed below.

In principle, it would be possible to calculate g^r using the eigenfunction expansion of the Green's function, Eq. (2.41). For that, we would need to calculate the eigenstates of $H_{\text{semi-infinite}}$. Such an eigenstate can in principle be constructed as a linear combination of solutions of the infinite wire. The coefficients of this linear combination must be chosen such that the boundary conditions of the isolated semi-infinite lead are fulfilled. This can be done easily if the transverse wave functions $\phi_{n,k}$ do not explicitly depend on k [59]: By combining propagating and counter-propagating modes, $\varphi(j) \sim \phi_n e^{ikj} - \phi_n e^{-ikj} \sim \phi_n \sin(kj)$, the wave function can be made zero at the end of the semi-infinite lead and thus satisfies the boundary condition there.

However, in general the transverse wave function $\phi_{n,k}$ can depend on k . For example, in a wire including a strong magnetic field the wave function propagating in one direction is localized at one edge of the wire, whereas the mode propagating in the opposite direction is localized at the opposite edge. In this case the construction of an eigenstate of the semi-infinite lead becomes very difficult and it is essential to include not only the propagating eigenstates $\varphi_{n,k}$, but also evanescent modes. The construction is even further complicated by the fact that different transverse wave functions $\phi_{n,k}$ and $\phi_{n',k'}$ are in general not orthogonal anymore, if $k \neq k'$. This is due to the fact that they are solutions to different effective Hamiltonians $H(k)$ and $H(k')$. Only the full wave functions $\varphi_{n,k}$ form an orthonormal set. Because of these difficulties we will not attempt to compute the eigenstates of the semi-infinite lead. Instead, we will first calculate the Green's function of the infinite wire for which the eigenstates are known and from that obtain the surface Green's function of the lead.

The problem of calculating the surface Green's function of a lead has been addressed frequently in the literature. First of all, the problem is strongly related to earlier accounts calculating the Green's function of an ideal crystal surface [94–97]. In contrast to a lead of finite transverse size as considered here, such a crystal surface is infinite and such are the respective Hamiltonian matrices. However, the problem can finally be reduced to a finite size invoking Bloch's theorem for the directions along the surface.

The numerical algorithms for computing the surface Green's function of a lead can be classified into two groups: Iterative methods and eigendecomposition based methods. The iterative algorithms calculate an approximate Green's function through some recursion relation [98–100] and require explicitly shifting the energy E by some small, but finite imaginary part. The eigendecomposition based methods compute the surface Green's function exactly—within the numerical precision—from the eigensolutions of an infinite wire. These methods have been derived independently several times [82, 86, 95–97, 101, 102], and equivalent expressions are also found in the mode matching method [103, 104], where instead of the surface Green's function the matching of lead eigenstates is considered¹. Compared to the iterative methods, the eigendecomposition based methods have been shown to be superior in accuracy and performance [101], and they provide additional information, such as the eigenmodes (channels) of the lead.

All of the accounts of the eigendecomposition based methods assume the hopping matrices H_1 , H_{-1} to be non-singular. However, singular—i.e. non-invertible, rank-deficient—hopping matrices arise naturally in important examples and are not at all unusual: For instance, a unit cell size larger than the range of the hopping in tight-binding approximation immediately leads to singular H_1 and H_{-1} [82, 86]. Graphene nanoribbons, subject of Chapters 6 and 7, are such an example. Iterative methods would be applicable to the case of singular hopping matrices, but their performance for singular hopping matrices is even worse than for non-singular hopping matrices. Still, in principle any problem involving singular hopping matrices could be recast into a problem involving only invertible hopping matrices [86]. However, the algorithm to do so is much more complicated than the expressions we will derive in the course of this chapter.

Thus, we will derive an eigendecomposition based expression for the surface Green's function equally valid for singular and non-singular hopping matrices H_1 , H_{-1} . To this end, the chapter is organized as follows: First, we will rigorously derive an expression for the surface Green's function of a lead analogous to previously derived expressions [82, 86, 101, 102], but without assuming the hopping matrices H_1 , H_{-1} to be invertible. We then discuss cases for which a direct implementation of this expression is numerically unstable and develop an algorithm circumventing these stability problems.

3.2 A general expression for the lead Green's function

3.2.1 Green's function of the infinite wire

Before we begin to calculate the Green's function of the infinite wire, we collect some useful relations that will allow us to extract the surface Green's function from that of

¹ The mode matching method is equivalent to the Green's function methods, as discussed in Ref. [105].

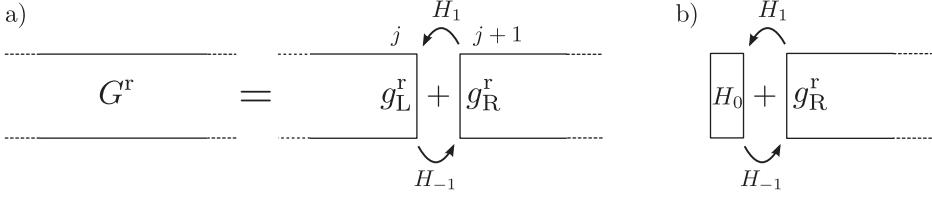


Figure 3.2 – (a) A infinite wire can be constructed by combining two semi-infinite wires.
(b) A semi-infinite wire does not change upon adding another unit cell.

the infinite wire in the subsequent sections.

The surface Green's function of the lead and the Green's function of the infinite wire can be related by the Dyson equation for the retarded Green's function, Eq. (2.91), in a similar fashion as in the recursive Green's function algorithm: The infinite wire can be regarded as a composition of two semi-infinite leads with the hopping between left and right lead as perturbation, as depicted in Fig. 3.2(a). We denote the retarded surface Green's function of the right (left) lead as g_R^r (g_L^r) and the retarded Green's function of the infinite wire as G^r . g_R^r is the desired surface Green's function of the lead. Note that because of translational invariance the surface Green's functions $g_{R,L}^r$ are independent of the position j , where the infinite wire is cut. Furthermore, g_L^r differs from g_R^r only therein, that the hopping Hamiltonians H_1 and H_{-1} are switched.

From the Dyson equation $G^r = G_0^r + G_0^r V G^r = G_0^r + G^r V G_0^r$, Eqs. (2.91) and (2.92), we can obtain relations between the surface Green's functions and the Green's function of the infinite wire. In the situation as depicted in Fig. 3.2(a), the unperturbed Green's function G_0^r is given by the surface Green's functions $g_{R,L}^r$ of the isolated leads and the perturbation V by the hopping matrices H_1 and H_{-1} . From this, we find:

$$G_{j,j'+1}^r = G_{j,j'}^r H_1 g_R^r \quad \text{for } j' \geq j \quad (3.8)$$

$$G_{j+1,j'}^r = g_R^r H_{-1} G_{j,j'}^r \quad \text{for } j' \leq j \quad (3.9)$$

Similar relations hold for g_L^r .

For the surface Green's function, another useful relation arises from the observation that the semi-infinite lead is invariant under addition of an extra unit cell (see Fig. 3.2(b)), and we obtain

$$g_R^r = (E - H_0 - H_1 g_R^r H_{-1})^{-1}. \quad (3.10)$$

Using the eigenstates of the infinite wire, Eq. (3.6), and the eigenstate representation of the retarded Green's function, Eq. (2.41), the retarded Green's function of the infinite

wire G^r is given as follows:

$$\begin{aligned}
G_{j,j'}^r &= \frac{1}{2\pi} \int_{-\pi}^{\pi} dk \sum_n \frac{\phi_{n,k} e^{ikj} \phi_{n,k}^\dagger e^{-ikj'}}{E - E_{n,k} + i\eta} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} dk \sum_n \underbrace{\left(\frac{1}{E - E_{n,k} + i\eta} \phi_{n,k} \right)}_{= \frac{1}{E - H(k) + i\eta} \phi_{n,k}} \phi_{n,k}^\dagger e^{ik(j-j')} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} dk \frac{1}{E - H(k) + i\eta} \underbrace{\left(\sum_n \phi_{n,k} \phi_{n,k}^\dagger \right)}_{= \mathbb{1}} e^{ik(j-j')} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} dk \frac{e^{ik(j-j')}}{E - H(k) + i\eta}, \tag{3.11}
\end{aligned}$$

where we have used that, for fixed k , the wave functions $\{\phi_{n,k}\}$ form a complete orthonormal set. By further substituting $z = e^{ik}$, the integration over k can be written as a contour integral on the unit circle and we arrive at

$$G_{j,j'}^r = \frac{1}{2\pi i} \oint dz \frac{z^{j-j'}}{(E + i\eta - H_0)z - H_1 z^2 - H_{-1}}. \tag{3.12}$$

It should be noted that the denominator of (3.12) is a matrix and the fraction should therefore be interpreted as a matrix inversion. Thus, the problem is more complicated than a scalar valued contour integral. Still, in order to solve the matrix-valued contour integral, we first identify the zeroes of the denominator, just as in the scalar case.

3.2.2 Zeroes of $(E + i\eta - \mathbf{H}_0)\mathbf{z} - \mathbf{H}_1\mathbf{z}^2 - \mathbf{H}_{-1}$

The contour integral of (3.12) is nonzero only if the entries of the matrix $((E + i\eta - H_0)z - H_1 z^2 - H_{-1})^{-1}$ have poles within the unit circle. The inverse of the matrix $\tilde{H}(E + i\eta, z) = (E + i\eta - H_0)z - H_1 z^2 - H_{-1}$ has a pole at λ_n , if the matrix $\tilde{H}(E + i\eta, z)$ itself is singular at $z = \lambda_n$. That is, we seek solutions of the (quadratic) eigenvalue equation

$$(E + i\eta - H_0)\lambda_n - H_1\lambda_n^2 - H_{-1} = 0. \tag{3.13}$$

In principle, the infinitesimal $i\eta$ could be taken care of in the numerical calculations by choosing some very small, but finite η . However, this introduces some unnecessary parameter dependence into the system, and η might need manual tuning in order to achieve convergence. We can avoid explicitly using η in the numerics by observing that η enters the calculation in a twofold way:

First, it determines which poles contribute to the contour integral in Eq. (3.12). In this respect, the infinitesimal η is crucial. Below, we will carefully distinguish different cases in order to account for this dependence on η .

Second, the values of the eigenvalues λ_n —i.e. the poles—and the eigenvectors \mathbf{u}_n depend on η : $\lambda_n = \lambda_n(E + i\eta)$, $\mathbf{u}_n = \mathbf{u}_n(E + i\eta)$. However, λ_n and \mathbf{u}_n are analytic functions of $E + i\eta$, and thus the dependence on η vanishes in the limit $\eta \rightarrow 0^+$. In particular, $\lim_{\eta \rightarrow 0^+} \lambda(E + i\eta) = \lambda_n(E)$, where $\lambda_n(E)$ is an eigenvalue of the quadratic eigenproblem *without* the infinitesimal shift $i\eta$. The same holds for the eigenvectors \mathbf{u}_n .²

Therefore, we seek solutions of the quadratic eigenvalue problem neglecting $i\eta$ (for a review on quadratic eigenproblems see [106])

$$((E - H_0)\lambda_n - H_1\lambda_n^2 - H_{-1}) \mathbf{u}_n = 0, \quad (3.14a)$$

or

$$\bar{\mathbf{u}}_n^\dagger ((E - H_0)\lambda_n - H_1\lambda_n^2 - H_{-1}) = 0, \quad (3.14b)$$

where \mathbf{u}_n ($\bar{\mathbf{u}}_n$) is a right (left) eigenvector for the eigenvalue λ_n . Such a quadratic eigenproblem has $2N_{\text{u.c.}}$ eigenvalues, when $N_{\text{u.c.}}$ is the size of the matrix $\tilde{H}(E, z)$.³ However, these eigenvalues are not necessarily finite, if H_{-1} is rank-deficient [106] and thus must be discussed carefully.

The quadratic eigenproblem can be recast into a linear eigenproblem, albeit of twice the size. By introducing the vector $\mathbf{v} = (\mathbf{u}, \lambda\mathbf{u})^t$ Eq. (3.14a) can be rewritten as

$$\begin{pmatrix} 0 & 1 \\ -H_{-1} & E - H_0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_i \\ \lambda_n \mathbf{u}_n \end{pmatrix} = \lambda_n \begin{pmatrix} 1 & 0 \\ 0 & H_1 \end{pmatrix} \begin{pmatrix} \mathbf{u}_n \\ \lambda_n \mathbf{u}_n \end{pmatrix}, \quad (3.15)$$

which is a generalized eigenproblem of the form $A\mathbf{v} = \lambda B\mathbf{v}$. Such a generalized eigenproblem can be recast into an ordinary eigenproblem if the matrix B is invertible, i.e. if the matrix H_{-1} is invertible:

$$\begin{pmatrix} 0 & 1 \\ -H_1^{-1}H_{-1} & H_1^{-1}(E - H_0) \end{pmatrix} \begin{pmatrix} \mathbf{u}_n \\ \lambda_n \mathbf{u}_n \end{pmatrix} = \lambda_n \begin{pmatrix} \mathbf{u}_n \\ \lambda_n \mathbf{u}_n \end{pmatrix}. \quad (3.16)$$

In any case, there are standard algorithms for solving both the ordinary and the generalized linear eigenproblem [107].

The eigenvalues of Eqs. (3.14a) and (3.14b) fall into two groups with $|\lambda_n| \neq 1$ and $|\lambda_n| = 1$. We first consider the case $|\lambda_n| \neq 1$. Then, λ_n lies inside the unit circle if $|\lambda_n| < 1$ and outside if $|\lambda_n| > 1$, regardless of the infinitesimal imaginary part of the energy $i\eta$. In general, λ_n can be written as $\lambda_n = e^{ik} = e^{i\Re(k)} e^{-i\Im(k)}$. $|\lambda_n| < 1$ corresponds to $\Im(k) > 0$ and an evanescent mode decaying to the right, whereas

² Note that the case of degenerate eigenvalues, $\lambda_n = \lambda_{n'}$ is a little more complicated in this respect and will be discussed in detail in the next section.

³ We can assume the quadratic eigenvalue problem to be regular. In an irregular eigenvalue problem $\det((E - H_0)\lambda - H_1\lambda - H_{-1}) = 0$ for all $\lambda \in \mathbb{C}$ which is unphysical. For example, in the infinite wire this would imply a flat band, $E(k) = \text{const.}$, with transverse wave function ϕ independent of k . This may only occur in ill-defined examples, such as a series of independent lattice points that are not coupled at all.

$|\lambda_n| > 1$ corresponds to $\Im(k) < 0$ and an evanescent mode decaying to the left. Furthermore, there is an equal number of right- and left-decaying solutions: By taking the complex conjugate of Eq. (3.14a) and using $H_1^\dagger = H_{-1}$, we find that for every eigenvalue λ_n , $1/\lambda_n^*$ is an eigenvalue of the quadratic eigenproblem, too [102]:

$$\mathbf{u}_n^\dagger ((E - H_0)1/\lambda_n^* - H_1(1/\lambda_n^*)^2 - H_{-1}) = 0. \quad (3.17)$$

Thus, for every eigenvalue within the unit circle there is a corresponding eigenvalue outside and vice versa.

As already mentioned above, the quadratic eigenvalue problem can have the eigenvalue ∞ , if the matrix H_{-1} (and thus also $H_1 = H_{-1}^\dagger$) is not invertible. In this case there is a vector \mathbf{u}_n such that $H_{-1}\mathbf{u}_n = 0$ and $\lambda_n = 0$ solves Eq. (3.14a). Then also $1/\lambda_n^*$ is an eigenvalue and corresponds to ∞ . It is therefore quite natural to obtain diverging eigenvalues if the matrix H_{-1} is rank-deficient and not at all problematic as opposed to the statements of Ref. [86]. The eigenvectors for $\lambda_n = 0, \infty$ itself do not have any physical meaning, however their contribution in terms of the contour integral in Eq. (3.12) is well-defined. Therefore, we also call $\lambda_n = 0$ (which lies within the unit circle) a “right-decaying” mode. The importance of these solutions has been first pointed out in the context of complex band structure calculations [108].

If $|\lambda_n| = |e^{ik}| = 1$, the eigenvectors \mathbf{u}_n are the propagating Bloch modes $\phi_{n,k}$ discussed above. Here we are considering a fixed energy E , and thus k is uniquely defined by $E_{n,k} = E$ and will be denoted by k_n . As a consequence, we will also drop the index k_n in the notation of the transverse wave function ϕ_n for the remainder of the chapter. Note that we will employ the notation ϕ_n if we want to emphasize that the eigenvector under consideration is a propagating mode, whereas \mathbf{u}_n denotes all eigenvectors, propagating and evanescent.

For the propagating modes ϕ_n the corresponding eigenvalues λ_n lie on the unit circle. Thus, the infinitesimal imaginary shift $i\eta$ becomes crucial in identifying whether the eigenvalue lies inside or outside of the unit circle. As shown by Taylor expanding $\lambda_n(E + i\eta)$ in appendix D.1, this is determined by the group velocity v_n of the Bloch mode ϕ_n .

The group velocity v_n is given by

$$\begin{aligned} v_n &= \frac{1}{\hbar} \left. \frac{dE_{n,k}}{dk} \right|_{k=k_n} \\ &= \frac{1}{\hbar} \left. \frac{d}{dk} \langle \phi_{n,k} | H_0 + H_1 e^{ik} + H_{-1} e^{-ik} | \phi_{n,k} \rangle \right|_{k=k_n}. \end{aligned} \quad (3.18)$$

Using the Feynman-Hellman theorem [109, 110] this expression can be simplified yield-

ing

$$\begin{aligned}
v_n &= \frac{1}{\hbar} \left\langle \phi_{n,k} \middle| \frac{d}{dk} (H_0 + H_1 e^{ik} + H_{-1} e^{-ik}) \right| \phi_{n,k} \rangle \Big|_{k=k_n} \\
&= \frac{i}{\hbar} \phi_n^\dagger (H_1 e^{ik_n} - H_{-1} e^{-ik_n}) \phi_n \\
&= -\frac{1}{\hbar} \Im \left(2 \phi_n^\dagger H_1 \lambda_n \phi_n \right). \tag{3.19}
\end{aligned}$$

As shown in appendix D.1, the eigenvalue λ_i lies inside the unit circle if $v_n > 0$, corresponding to a right-moving wave, and outside if $v_n < 0$, corresponding to a left-moving wave.

Note that in this situation we cannot invoke Eq. (3.17) to make a statement about the number of right- and left-moving modes, as $1/\lambda_n^* = 1/(e^{ik})^* = e^{ik} = \lambda_n$. Eq. (3.17) thus only shows that for any right eigenvector ϕ_n , the Hermitian conjugate ϕ_n^\dagger is a left eigenvector to the same eigenvalue λ_n . Still, we can determine the number of right- and left-moving waves by considering the properties of the Bloch bands $E_{n,k}$. The Bloch bands $E_{n,k}$ are periodic in k with periodicity 2π . Therefore, at a given energy E , there is an equal number of right- and left-moving waves, as the Bloch bands $E_{n,k}$ intersect the energy E equally often with positive and negative slope [82].

Summarizing the arguments of this section, we find that the eigenvalues and -vectors of the quadratic eigenproblem, Eq. (3.14a), can be grouped into

- $N_{u.c.}$ right-going modes with $|\lambda_i| < 1$ or $|\lambda_i| = 1 \wedge v_n > 0$. These eigenvalues are denoted by $\lambda_{i,<}$, the corresponding eigenvectors by $\mathbf{u}_{i,<}$ or $\phi_{n,<}$. In addition, we define the matrix of right-going modes as $U_< = (\mathbf{u}_{1,<} \dots, \mathbf{u}_{N_{u.c.},<})$.
- $N_{u.c.}$ left-going modes with $|\lambda_i| > 1$ or $|\lambda_i| = 1 \wedge v_n < 0$. These eigenvalues are denoted by $\lambda_{i,>}$, the corresponding eigenvectors by $\mathbf{u}_{i,>}$ or $\phi_{n,>}$. In addition, we define the matrix of left-going modes as $U_> = (\mathbf{u}_{1,>} \dots, \mathbf{u}_{N_{u.c.},>})$.

It should be emphasized that these eigenvalues and -vectors can be computed and classified without having to resort to a finite imaginary shift $i\eta$.

3.2.3 The role of degenerate modes

The solutions of the quadratic eigenvalue problem (3.14a) are only unique, if all eigenvalues λ_n are different. If \mathbf{u}_n and $\mathbf{u}_{n'}$ are solutions with the same eigenvalue $\lambda = \lambda_n = \lambda_{n'}$, then any linear combination solves the quadratic eigenproblem:

$$((E - H_0) \lambda + H_1 \lambda^2 + H_{-1}) (a \mathbf{u}_n + a' \mathbf{u}_{n'}) = 0 \quad \text{for } a, a' \in \mathbb{C}. \tag{3.20}$$

Although this does not pose a problem for decaying solutions⁴, it very well can for the propagating solutions, where $\lambda_n = e^{ik_n}$. As shown in Section 3.2.2, the group velocity

⁴ Decaying solutions are classified only according to their eigenvalue λ_n , the eigenvector \mathbf{u}_n does not enter. Furthermore, the final results, given in Eqs. (3.28) and (3.29) are invariant under a change of basis, i.e. choosing different linear combinations of degenerate solutions.

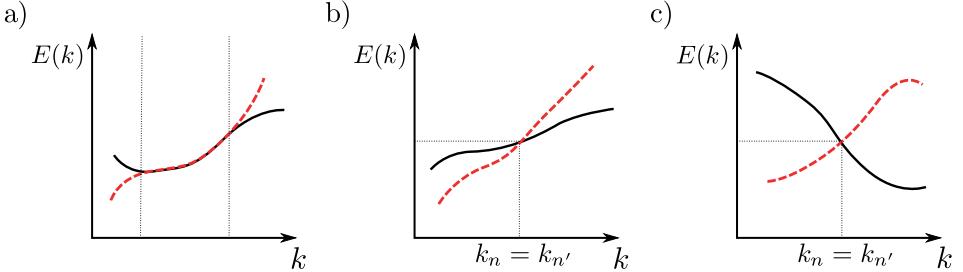


Figure 3.3 – Schematic band structure plots for different cases of degeneracies in Bloch modes: (a) Two bands are degenerate for a whole range of k -values. (b) Two bands cross at a single k -point with slopes of the same sign. (c) Two bands cross with slopes of opposite signs.

v_n of propagating modes determines whether a solution lies inside or outside the unit circle. If two or more propagating modes share the same eigenvalue, i.e. $k_n = k_{n'}$, the group velocity becomes ill-defined, as different linear combinations may result in different group velocities. Moreover, the eigensolutions $\phi_n(E)$ are not guaranteed to be smooth functions around $E = E_{n,k_n}$: Whereas at the degeneracy point E_{n,k_n} any linear combination of degenerate modes is a solution, moving away from this point in the complex plane will lift the degeneracy and lead to definite solutions.⁵

The different types of degeneracies that can occur are depicted schematically in Fig. 3.3. First, energy bands can be degenerate over a range of k -values, as shown in Fig. 3.3(a). Such a degeneracy can be due to symmetries of the underlying lattice, or occurs naturally if the system is degenerate with respect to an additional quantum number, such as spin. Second, two or more bands can intersect with a slope (and thus velocity) of the same sign (Fig. 3.3(b)). Both cases are not problematic, as then all degenerate eigenvalues are either inside or outside the unit circle simultaneously. If however, as in the third case (Fig. 3.3(c)), bands intersect with slopes of opposite sign, some of the degenerate eigenvalues are located inside and some outside the unit circle.

The troublesome degeneracy of eigenstates is an artifact of neglecting the infinitesimal shift $i\eta$. We can however find a remedy by observing that for non-degenerate λ_n the eigenvectors ϕ_n diagonalize the velocity operator (Eq. (3.19)), as shown in appendix C; this is not necessarily the case for degenerate eigenvalues. We can expect properties that depend on the eigenvectors ϕ_n , such as the diagonal form of the velocity operator, to change in a continuous way. Therefore, the condition that the velocity operator should be diagonal with respect to the eigenvectors ϕ_n uniquely defines a set of basis vectors with definite group velocity v_n that can be used to reliably classify these modes

⁵ Indeed, the smoothness of $\phi_n(E)$ is a prerequisite of the derivation in appendix D.1 that shows that the group velocity of propagating modes determines whether a eigenvalue lies inside or outside the unit circle.

as right- and left-moving.

Suppose that $\phi_{n_1}, \dots, \phi_{n_l}$ are eigenvectors for the eigenvalue $\lambda = e^{ik}$. Let $\Phi = (\phi_{n_1}, \dots, \phi_{n_l})$ be the matrix built from these eigenvectors. Then, the velocity operator in the subspace of these eigenvectors can be written as

$$V_\Phi = \Phi^\dagger i(H_1 e^{ik} - H_{-1} e^{-ik})\Phi. \quad (3.21)$$

V_Φ is Hermitian and thus can be diagonalized [82] as $V'_\Phi = S^\dagger V_\Phi S$ with some unitary matrix S . From the columns of $\Phi' = \Phi S$ we then find the new eigenvectors $\phi'_{n_1}, \dots, \phi'_{n_l}$ with well-defined velocities. Furthermore, the rotated ϕ'_n are smooth functions with respect to energy E .

3.2.4 Solving the contour integral

Before proceeding further in solving the contour integral in Eq. (3.12), we first study the eigenvalues and -vectors of $g_R^r H_{-1}$, as this can give some indication on the final expression for the surface Green's function. Suppose \mathbf{u} is an eigenvector of $g_R^r H_{-1}$ with eigenvalue λ :

$$g_R^r H_{-1} \mathbf{u} = \lambda \mathbf{u}. \quad (3.22)$$

Then we find

$$\begin{aligned} H_{-1} \mathbf{u} &= (g_R^r)^{-1} \lambda \mathbf{u} \\ &= (E - H_0 - H_1 g_R^r H_{-1}) \lambda \mathbf{u} \\ &= ((E - H_0) \lambda - H_1 \lambda^2) \mathbf{u}, \end{aligned} \quad (3.23)$$

where we have used Eqs. (3.10) and (3.22). Every eigenvalue λ is thus also a solution of the quadratic eigenproblem

$$((E - H_0) \lambda - H_1 \lambda^2 - H_{-1}) \mathbf{u} = 0. \quad (3.24)$$

$g_R^r H_{-1}$ can be expressed in terms of its eigenvectors and -values and therefore takes the form

$$g_R^r H_{-1} = U \Lambda U^{-1}, \quad (3.25)$$

where the matrices U and Λ are built up by choosing the appropriate eigenvectors and -values from the $2N$ solutions of the quadratic eigenproblem (3.14a). Unfortunately, these considerations do not give an indication on *which* solutions should be chosen.

The appropriate eigenvalues and -vectors can be selected by using the boundary condition of the retarded Green's function, namely that it should contain only outgoing waves [59, 82]. However, this intuitive physical arguments break down if systems with rank-deficient matrices H_1 and H_{-1} are considered and, as argued above, spurious solutions with vanishing or infinite eigenvalues appear. These solutions do not have any physical meaning, but are well-defined in the context of the contour integral of Eq. (3.12). Therefore we resort to this contour integration to extend the previous

expressions for the surface Green's function [82, 86, 101–103] also to the important case of rank-deficient hopping matrices.

We start from Eq. (3.12) and insert $U_< (U_<)^{-1} = \mathbb{1}$:⁶

$$\begin{aligned} G_{j,j'}^r &= \frac{1}{2\pi i} \oint dz \frac{z^{j-j'}}{(E + i\eta - H_0)z - H_1 z^2 - H_{-1}} \\ &= \frac{1}{2\pi i} \oint dz U_< (U_<)^{-1} \frac{z^{j-j'}}{\tilde{H}(E + i\eta, z)} \\ &= \frac{1}{2\pi i} \oint dz U_< \frac{z^{j-j'}}{\tilde{H}(E + i\eta, z) U_<}. \end{aligned} \quad (3.26)$$

The denominator $\tilde{H}(E + i\eta, z)U_<$ now has columns $\tilde{H}(E + i\eta, z)\mathbf{u}_{i,<}$ with zeroes at $\lambda_{i,<}$. By carefully counting the multiplicities of the zeroes in the numerator and the denominator of the inverse $(\tilde{H}(E + i\eta, z)U_<)^{-1}$, we can show that (see appendix D.2)

$$\begin{aligned} G_{j,j'}^r &= \frac{1}{2\pi i} \oint dz U_< \Lambda_{<}^{j-j'} \frac{1}{\tilde{H}(E + i\eta, z) U_<} \\ &= U_< \Lambda_{<}^{j-j'} U_<^{-1} \frac{1}{2\pi i} \oint dz \frac{1}{(E + i\eta - H_0)z - H_1 z^2 - H_{-1}} \\ &= U_< \Lambda_{<}^{j-j'} U_<^{-1} G_{j',j'}^r \quad \text{for } j' \leq j. \end{aligned} \quad (3.27)$$

Setting $j = j' + 1$ and using Eq. (3.9), $G_{j'+1,j'}^r = g_R^r H_{-1} G_{j',j'}^r$, we then obtain the expression:⁷

$$g_R^r H_{-1} = U_< \Lambda_{<} U_<^{-1} \quad (3.28)$$

valid also for singular H_{-1} . This is the main finding of this chapter.

From Eq. (3.28) we immediately obtain an expression for the self-energy of the lead⁸, $\Sigma_R = H_1 g_R^r H_{-1} = H_1 U_< \Lambda_{<} U_<^{-1}$. If the surface Green's function g_R^r itself is needed, it can be calculated either by multiplying Eq. (3.28) with $(H_{-1})^{-1}$ from the right, if H_{-1} is invertible, or through Eq. (3.10), which is also valid for rank-deficient hopping matrices.

⁶ Here we have to assume without proof that $U_<$ is invertible. It is not obvious that the $N_{\text{u.c.}}$ right-going solutions $\mathbf{u}_{i,<}$ of the quadratic eigenproblem (3.14a) are linearly independent: For example, we have already seen that they are not necessarily orthogonal, as they are solutions to different Hamiltonians $H(\lambda_{i,<})$.

The eigenvectors of the linearized eigenproblem, $(\mathbf{u}_{i,<}, \lambda_{i,<} \mathbf{u}_{i,<})^t$ do form a linearly independent set, as they are solutions to a (generalized) linear eigenproblem. Unfortunately, this by itself does not prove the linear independence of the $\mathbf{u}_{i,<}$ alone.

⁷ This is uniquely defined, as $G_{j',j'}^r = (E - H_0 - H_1 g_R^r H_{-1} - H_{-1} g_L^r H_1)^{-1}$ is invertible.

⁸ Here, we assume that the hopping from lead l to the scattering region, H_{Sl} , is given by H_1 . Such a situation can always be achieved by formally assigning the first unit cell of the lead to the scattering region.

In an analogous way we can show that every left eigenvector $\bar{\mathbf{u}}$ of $H_1 g_R^r$ with eigenvalue λ is a left eigenvector of the quadratic eigenvalue equation with eigenvalue $1/\lambda$. Consequently, we find that $H_1 g_R^r$ can be written in terms of left-going solutions:

$$H_1 g_R^r = \left(\bar{U}_>^\dagger \right)^{-1} \Lambda_>^{-1} \bar{U}_>^\dagger. \quad (3.29)$$

If we were to calculate the advanced surface Green's function g_R^a , we would choose in Eq. (3.12) a negative infinitesimal shift, $E - i\eta$. The subsequent calculation proceeds in the same fashion as for the retarded Green's function, except that the propagating modes are handled differently: $-i\eta$ now selects *left-moving* instead of right-moving propagating modes. In contrast, the decaying modes are unchanged with respect to the retarded case.

We have found expressions for the surface Green's function in terms of eigenstates of the infinite wire. It should be emphasized that these expressions are fundamentally different from the normal eigenstate representation of the Green's function, Eq. (2.41). In the usual eigenstate representation, the Green's function is expressed in terms of *all* eigenstates of the system, and thus *all* energies $E_{n,k}$. In contrast, the expressions derived in this chapter express the surface Green's function in terms of eigenstates for a *single* energy E . The Green's function is in this case not only expressed in terms of physical eigenstates of the (infinite) system, but also unphysical ones: The decaying eigenstates grow exponentially in one direction and thus cannot be normalized. The fact that the eigenstates for a single energy E are enough to calculate the surface Green's function make these expressions very useful for numerical computations. A summary of the algorithm based on the results of this section can be found in Appendix D.3.1.

3.3 A more stable algorithm based on the Schur decomposition

3.3.1 Failure of the eigendecomposition based algorithm

The expressions for the surface Green's function of a lead are well-suited for a numerical implementation. However, it turns out that numerics based on these expressions can fail for certain, important cases. In order to illustrate this failure and to trace its origin, we will consider the example of an electronic wire in a magnetic field. The lessons learned from this example will then serve as guidelines for developing a more stable algorithm.

We will consider the case of a two-dimensional electron gas with a confinement potential of the form $V(x, y) = V(y)$, where

$$V(y) = \begin{cases} 0 & \text{for } 0 < y < W, \\ \infty & \text{else,} \end{cases} \quad (3.30)$$

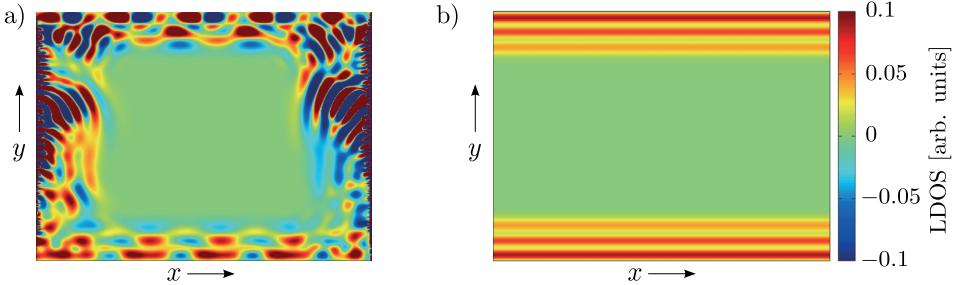


Figure 3.4 – Local density of states of wire in a perpendicular magnet field B , obtained numerically using the lead Green's functions calculated with (a) the eigendecomposition based algorithms of Section D.3.1 and (b) the Schur decomposition based algorithms of Section D.3.2. The parameters used were $W/l_{\text{cycl}} = 9$, $W = 150a$ and $E = 0.15 \frac{\hbar^2}{2ma_0^2}$, where a_0 is the lattice constant of the discretization and m the effective electron mass.

i.e. the electron gas is restricted to a wire of width W . Furthermore we assume a homogeneous magnetic field B perpendicular to the plane of the electron gas. Such a geometry is a simplified version of the systems considered e.g. for weak localization or quantum Hall effect, and as such reliable numerical solutions are of importance.

The Hamiltonian of the system is given by

$$H = \frac{1}{2m} (-i\hbar\nabla - e\mathbf{A}(\mathbf{x}))^2 + V(y), \quad (3.31)$$

where m is the effective mass of the system and e the electronic charge. We choose the vector potential $\mathbf{A}(\mathbf{x})$ in x -direction, $A_y = 0$, $A_x(y) = -By$. After discretizing the Hamiltonian (for details, see Appendix F), the Hamiltonian matrix of the unit cell H_0 is independent of the magnetic field. The magnetic field dependence is contained in the hopping matrices: The hopping matrix H_1 is a diagonal matrix with entries $(H_1)_{nn} = e^{ian}$, where α is some phase related to the magnetic field B and the lattice constant of the discretization. Thus, H_1 is invertible ($(H_1)^{-1} = H_1^\dagger = H_{-1}$) and we can use the expressions for invertible hopping matrices. We will measure the strength of the magnetic field in terms of the cyclotron radius l_{cycl} with respect to the wire width W , where $l_{\text{cycl}} = \frac{p_F}{eB}$ [111], with p_F the Fermi momentum.

Wires are a rather useful tool for testing numerical transport algorithms: Due to the translational symmetry in x direction, observables must be constant along the wire direction. Any spatial dependence along the wire is thus a clear indication of numerical errors. Furthermore, transport in large magnetic fields, i.e. in the quantum Hall regime, is very convenient for checking numerical calculations, as the wave functions of the edge states are relatively simple. Programming errors or failures of the numerics can then often be detected by visual inspection of observables, such as the local density of states.

In Fig. 3.4 we show the LDOS of a wire in magnetic field. The LDOS was calculated using the recursive Green's function technique first with the lead Green's function cal-

culated from the expressions of the previous section (Fig. 3.4(a)), and second with the lead Green's function calculated with the stable algorithm developed in the remainder of the chapter (Fig. 3.4(b)). Fig. 3.4(b) shows nicely three edge modes forming a positive LDOS that is constant in x -direction, and we can consider this solution to be exact within the numerical limits. In contrast, Fig. 3.4(a) shows some complex spatial behavior of the LDOS with positive and even negative values, violating the simple physical considerations for wires mentioned above. When moving away from the leads, further into the region calculated via the RGF algorithm, some of the unphysical states introduced by the incorrect lead Green's function decay and the LDOS starts to resemble the exact solution. In fact, if we make the region calculated via the RGF algorithm longer, eventually all the unphysical solutions will decay so much that the LDOS will finally start to resemble the exact solution. This is a clear indication that indeed the problem lies in the (failed) calculation of the lead Green's function. Obviously this failure has catastrophic consequences here.

As we will show below, the failure of the algorithm of Section D.3.1 can be attributed to the inversion of $U_<$, that becomes ill-conditioned for the problem under discussion. To this end, we first define some mathematical measure in order to quantify how well-posed the problem of inversion is numerically. The *condition number* of a matrix A is defined as

$$\text{cond}(A) = \|A\|_2 \cdot \|A^{-1}\|_2, \quad (3.32)$$

where $\|\dots\|_2$ is the matrix 2-norm [112]. The numerical error in inverting a matrix A can then be estimated as $\approx \varepsilon \text{cond}(A)$, where ε is the machine precision. On a typical computer, the standard floating point data type is an eight byte `double` with $\varepsilon \approx 10^{-16}$. Numerically, the condition number can be estimated from the singular value decomposition of A as the ratio of the largest and the smallest singular value [112].

Another useful measure in this context is the rank of matrix A , $\text{rank}(A)$. If the columns (or rows) of a $m \times n$ -matrix A form a set of linearly independent vectors, the matrix has full rank, i.e. $\text{rank}(A) = \min(m, n)$. A square matrix A is invertible iff it has full rank. Numerically, we can calculate the *approximate rank* of a matrix A by counting the non-zero singular values of A [112]. This rank is only approximate as we need to specify some numerical tolerance in order to check whether a numerically computed singular value is zero, and the rank will depend on this tolerance. Still, the approximate rank can quantify whether the columns (or rows) of a matrix are linearly dependent or at least close to linearly dependent. The relation of the condition number and the approximate rank is obvious from the way they are calculated numerically: The condition number will become large, if at least a single singular value is close to zero, i.e. if a matrix has not full rank. In return, the approximate rank depends on the total number of singular values close to zero and thus gives a further estimate on the character of the ill-conditioned problem.

In Fig. 3.5(a) we show the condition number of $U_<$ as a function of $W/l_{\text{cycl}} \propto B$. For small magnetic field, the condition number of $U_<$ is small, indicating a well-conditioned problem. When the magnetic field increases and W/l_{cycl} approaches 1—

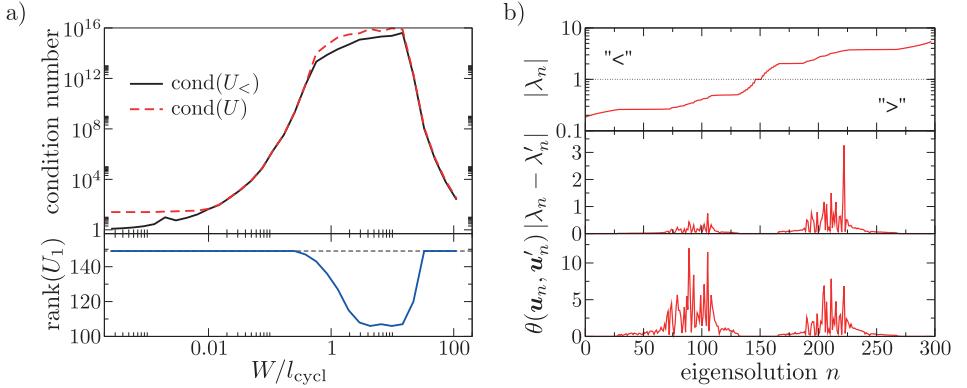


Figure 3.5 – (a) condition numbers (upper panel) and approximate $\text{rank}(U_1)$ (lower panel) for the eigendecomposition based algorithm as a function of magnetic field strength W/l_{cycl} . The dashed line in the lower panel shows the value $\text{rank}(U_1)$ for linearly independent eigenvectors. (b) Detailed analysis of error bounds of individual eigenvalues λ_n for fixed magnetic field, $W/l_{\text{cycl}} = 9$: absolute value of λ_n (upper panel), error bound to the eigenvalue λ_n (middle panel), and error bound to the eigenvector \mathbf{u}_n (lower panel) as a function of eigenvalue index n . The eigenvalues and -vectors have been sorted such that the right-going set (“<”) precedes the left-going set (“>”). The calculations were done using the same set of parameters as in Fig. 3.4.

i.e. the magnetic field begins to influence the wave functions of the wire beyond simple perturbations— $\text{cond}(U_<)$ takes values in the range of $10^{10} - 10^{16}$. Numerically, the inversion of $U_<$ becomes meaningless in this case and so does the calculated surface Green's function. Only for very large magnetic field the matrix $U_<$ becomes well-conditioned again. The magnetic field range for which $U_<$ is ill-conditioned increases as the width W of the system becomes larger, i.e. the numerical error increases as the system size gets larger. This is an indication that the inversion of $U_<$ becomes ill-defined due to numerical errors, rather than $U_<$ itself being non-invertible.

The $N_{\text{u.c.}} \times N_{\text{u.c.}}$ matrix $U_<$ is constructed from selected eigenvectors $(\mathbf{u}_n, \mathbf{u}_n \lambda_n)^t$ of the $2N_{\text{u.c.}} \times 2N_{\text{u.c.}}$ linearization of the quadratic eigenvalue problem, Eq. (3.14a). The full matrix U of all eigenvectors $(\mathbf{u}_n, \mathbf{u}_n \lambda_n)^t$ can be written as

$$U = \begin{pmatrix} U_< & U_> \\ U_< \Lambda_< & U_> \Lambda_> \end{pmatrix} = (U_1 \quad U_2) . \quad (3.33)$$

In Fig. 3.5(a) we also show that $\text{cond}(U) \approx \text{cond}(U_<)$, in order to demonstrate that the failure of numerics is not due to choosing the submatrix $U_<$, but is inherent to the linearized quadratic eigenproblem itself. This can also be seen from the approximate $\text{rank}(U_1)$ in the lower panel of Fig. 3.5(a). Whereas the rank of U_1 should be $N_{\text{u.c.}}$, as the $N_{\text{u.c.}}$ eigenvectors span a $N_{\text{u.c.}}$ -dimensional subspace, the numerically computed U_1 is rather rank-deficient in the magnetic field range where the numerics breaks down.

It should be noted that the linearized (generalized) eigenproblem of Eq. (3.15) involves computing the eigenvalues and -vectors of *non-Hermitian* matrices. In contrast to Hermitian matrices, the eigenvalues and -vectors of non-Hermitian matrices can be rather ill-conditioned [112]. For example, the eigenvectors of a non-Hermitian matrix need not be orthogonal, and thus the matrix U , while in principle invertible, could be easily near-singular.

We therefore analyze the error bounds of the numerically computed eigenvalues and -vectors. Suppose that λ'_n and \mathbf{u}'_n is the exact eigensolution. Then the error in the numerically computed eigenvalue λ_n and -vector \mathbf{u}_n can be quantified as $|\lambda_n - \lambda'_n|$ and the acute angle $\theta(\mathbf{u}_n, \mathbf{u}'_n)$, respectively [107]. In Fig. 3.5(b) we show the errors on the n -th eigenvalues and -vector together with the absolute value of the eigenvalue $|\lambda_n|$ for all $2N_{\text{u.c.}}$ eigensolutions in the regime where $U_<$ is ill-conditioned ($W/l_{\text{cycl}} = 9$, corresponding to the density plot in Fig. 3.4(a)). We find two clusters of eigensolutions in both the right- and left-going solutions with large error bounds: The error in the eigenvalues is of the order of the eigenvalue itself, and the angles between numerical and exact eigenvectors are of the order of π . Obviously, these solutions are not reliable. Note that these numerical problems are restricted to decaying solutions, whereas the propagating solutions (with $|\lambda_n| = 1$) have small error bounds. Presumably this is due to the fact that the propagating modes are solutions to a Hermitian Hamiltonian $H(k)$.

Summarizing the discussion, the failure of the eigendecomposition based algorithms for computing lead Green's functions is due to numerical errors in computing the eigenvalues and -vectors of a non-Hermitian matrix. This leads numerically to linearly dependent eigenvectors and thus to non-invertible eigenvector matrices.

Having established the cause of the failure of numerics as numerical errors due to finite precision, we are left with two options: First, we could resort to floating point arithmetics with higher precision. However, there are no readily available numerical routines beyond double precision, and extending the established linear algebra routines to higher precision is a major task and not trivial. Moreover, it is not clear whether higher precision would solve all problems. The condition numbers of Fig. 3.5(a) were calculated numerically. Thus they are also limited by the floating point precision and can only give a lower bound—the true condition number could be much larger. Due to these difficulties associated with higher precision we choose the second option: Develop an alternative algorithm that does not need to calculate all individual eigenvectors and -values.

3.3.2 Surface Green's function from the Schur decomposition

In order to develop a more stable algorithm for calculating the surface Green's function of a lead, we introduce the concept of an *invariant subspace* [112]. An invariant subspace of a matrix A is a vector space V such that for every $v \in V$, also $Av \in V$. For example, a subspace spanned by eigenvectors of A is an invariant subspace.

Starting point for a more stable algorithm is the observation that it is indeed not necessary to calculate individual eigensolutions, but only the invariant subspace associated

with a cluster of eigenvalues. The columns of the eigenvector matrix U_1 ,

$$U_1 = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} = \begin{pmatrix} U_< \\ U_<\Lambda_< \end{pmatrix}, \quad (3.34)$$

form an invariant subspace $\text{range}(U_1)$. Any invertible $N_{\text{u.c.}} \times N_{\text{u.c.}}$ -matrix A generates a basis transformation of this subspace and vice versa, so that the columns of $U'_1 = U_1 A$ span the same subspace, $\text{range}(U'_1) = \text{range}(U_1)$. Then we find that

$$\begin{aligned} U'_{21} (U'_{11})^{-1} &= U_{21} A (U_{11} A)^{-1} \\ &= U_<\Lambda_< A A^{-1} U_<^{-1} \\ &= U_<\Lambda_< U_<^{-1}. \end{aligned} \quad (3.35)$$

Therefore, it is enough to compute the invariant subspace associated with the eigenvalues $\lambda_{n,<}$ in order to compute the surface Green's function of the lead.

In the example of the previous section, the failure of the numerics was linked to a rank-deficient matrix U_1 , i.e. a too small invariant subspace. Since the eigenvectors of a non-Hermitian matrix do in general not form an orthogonal basis, the eigenvectors can be nearly linearly dependent and numerical errors may lead to rank-deficient eigenvector matrices. Therefore, we seek an alternative method to compute the invariant subspace spanned by U_1 .

We start with discussing the case of invertible hopping matrices. The linearized quadratic eigenproblem can in this case be written as an ordinary eigenproblem of the form $A(\mathbf{u}_n, \lambda_n \mathbf{u}_n)^t = \lambda_n (\mathbf{u}_n, \lambda_n \mathbf{u}_n)^t$, as in Eq. (3.16). Then, there is a unitary transformation

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = (Q_1 \quad Q_2) \quad (3.36)$$

such that

$$Q^\dagger A Q = T = \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix}, \quad (3.37)$$

where T is an upper triangular matrix with the eigenvalues λ_n of A on the diagonal. This decomposition is called *Schur decomposition*. Furthermore, the unitary transformation Q can be chosen such that the eigenvalues λ_n appear in any desired order along the diagonal. The first M columns of Q , $1 \leq M \leq 2N_{\text{u.c.}}$, then form an invariant subspace with respect to the first M eigenvalues, $\lambda_1 \dots \lambda_M$ [112].

Suppose now that the eigenvalues are ordered such that all eigenvalues $\lambda_{n,<}$ are located in the upper left $N \times N$ -block T_{11} . If T_{11} and T_{22} do not share common eigenvalues, the triangular matrix T can be block-diagonalized [112], i.e. there is a non-singular transformation

$$Y = \begin{pmatrix} \mathbb{1} & Y_{12} \\ 0 & \mathbb{1} \end{pmatrix} \quad (3.38)$$

such that

$$Y^{-1} T Y = Y^{-1} Q^\dagger A Q Y = \begin{pmatrix} T_{11} & 0 \\ 0 & T_{22} \end{pmatrix}. \quad (3.39)$$

After this transformation, the $2N_{\text{u.c.}} \times 2N_{\text{u.c.}}$ eigenproblem is reduced to two separate $N_{\text{u.c.}} \times N_{\text{u.c.}}$ eigenproblems for the eigenvalues $\lambda_{n,<}$ and $\lambda_{n,>}$, respectively. The matrix A can thus be reduced to diagonal form—i.e. the full eigenproblem can be solved—by a block-diagonal transformation

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \quad (3.40)$$

such that

$$X^{-1}Y^{-1}Q^\dagger A Q Y X = \begin{pmatrix} \Lambda_{<} & 0 \\ 0 & \Lambda_{>} \end{pmatrix}. \quad (3.41)$$

The matrix of eigenvectors is then given by $U = QXY$. Hence, we find

$$\begin{aligned} g_R^r H_{-1} &= U_{<} \Lambda_{<} U_{<}^{-1} = Q_{21} X_1 (Q_{11} X_1)^{-1} \\ &= Q_{21} Q_{11}^{-1}. \end{aligned} \quad (3.42)$$

The surface Green's function can thus be calculated from the *orthonormal* basis vectors of the invariant subspace associated with $\lambda_{n,<}$ obtained from the Schur decomposition, instead of the full eigendecomposition.

It should be noted that in the derivation of Eq. (3.42) it is necessary to assume that T_{11} and T_{22} do not share any eigenvalues. Hence, Eq. (3.42) is not valid in a situation as depicted in Fig. 3.3(c), when propagating modes share the same eigenvalue $\lambda_n = \lambda_{n'}$, but have group velocities v_n and $v_{n'}$ with opposite signs. In such a case it is necessary to explicitly shift the real energy either along the real or the imaginary axis (i.e. with a finite $i\eta$), lifting the eigenvalue degeneracy in order to apply Eq. (3.42). In any case, it is rather unlikely to hit such a situation numerically.

Finally, we turn to the case of a singular hopping matrix H_1 . The linearized quadratic eigenproblem forms in this case a generalized eigenvalue problem $A(\mathbf{u}_n, \lambda_n \mathbf{u}_n)^t = \lambda_n B(\mathbf{u}_n, \lambda_n \mathbf{u}_n)^t$, as in Eq. (3.15). Then, there are two unitary transformations Q and Z such that

$$Q^\dagger AZ = T \quad \text{and} \quad Q^\dagger BZ = S, \quad (3.43)$$

where T and S are upper triangular [112] (*generalized Schur decomposition*). The eigenvalues of the generalized eigenproblem are thence given by $\lambda_n = t_{nn}/s_{nn}$, where t_{nn} (s_{nn}) are the diagonal elements of T (S). Again, we assume that the eigenvalues are ordered such that the $\lambda_{n,<}$ are contained in the leading diagonal positions of T and S . In a similar fashion as in the case of invertible hopping matrices, we find that the Green's function can be expressed in terms of the subspace spanned by the first $N_{\text{u.c.}}$ orthonormal columns of Z :

$$g_R^r H_{-1} = Z_{21} Z_{11}^{-1}. \quad (3.44)$$

A summary of the numerical algorithms based on the results of this section can be found in Appendix D.3.2.

3.3.3 Numerical Stability

The Schur decomposition based algorithm developed in the previous sections is superior to the standard, eigendecomposition based algorithms in terms of numerical stability. Although we will not proof this statement in a strict, mathematical sense, we will give some arguments why this is the case below.

In Section 3.3.1, the failure of the standard algorithm was linked to the rank-deficiency of the matrix U_1 . The numerically computed eigenvectors did not form a linearly independent set, and thus did not span the full invariant subspace $\text{range}(U_1)$ associated with the eigenvalues $\lambda_{n,<}$.

In contrast to the eigendecomposition of a non-Hermitian matrix, the (generalized) Schur decomposition can be computed more stably, as it involves only unitary transformations [112]. In addition, reordering the (generalized) Schur form can be done in a numerically stable fashion, too. Since the transformation matrix Q/Z is unitary, its columns form an orthonormal basis set, as opposed to the basis spanned by the eigenvectors. Therefore, the Schur decomposition yields an orthonormal basis set for the invariant subspace $\text{range}(U_1) = \text{range}(Q_1/Z_1)$. Due to the orthogonality of the N basis vectors, even when obtained numerically, the computed subspace will *always* have dimensionality N , in contrast to the standard algorithm.

Apart from the correct dimensionality, the computed invariant subspace must be associated with the correct eigenvalues. Note that we need not determine every single eigenvalue accurately, but we only need to cluster the eigenvalues into left- and right-going sets $\{\lambda_{n,>}\}$ and $\{\lambda_{n,<}\}$. The border of these sets is within the propagating modes where $|\lambda_n| = 1$. In the examples considered in this study, we have found that the error bounds on the eigenvalues and -vectors around $|\lambda_n| = 1$ are rather small (for an example, see Fig. 3.5(b)). Presumably this is due to the fact that the propagating modes (i.e. $|\lambda_n| = 1$) are governed by a Hermitian operator. Hence, the eigenvalues and -vectors separating left- and right-going sets can be calculated accurately and we can thus expect to obtain an accurate clustering. Indeed, the examples in the next section will show that we can calculate the invariant subspace $\text{range}(U_1) = \text{range}(Q_1/Z_1)$ very accurately.

Finally, after calculating the invariant subspace we need to invert the block Q_{11}/Z_{11} . Unfortunately we cannot prove that this matrix is always well-conditioned—whereas the full matrix Q/Z is unitary and thus well-conditioned, the subblock Q_{11}/Z_{11} is in general not unitary. However, we found that it was always well-behaved in our examples, as shown in the next section.

In summary, while eigendecomposition and Schur decomposition based algorithms need to perform several computational steps alike, such as the classification of eigenvalues, the Schur decomposition performs significantly better in one crucial step, the computation of the invariant subspace.

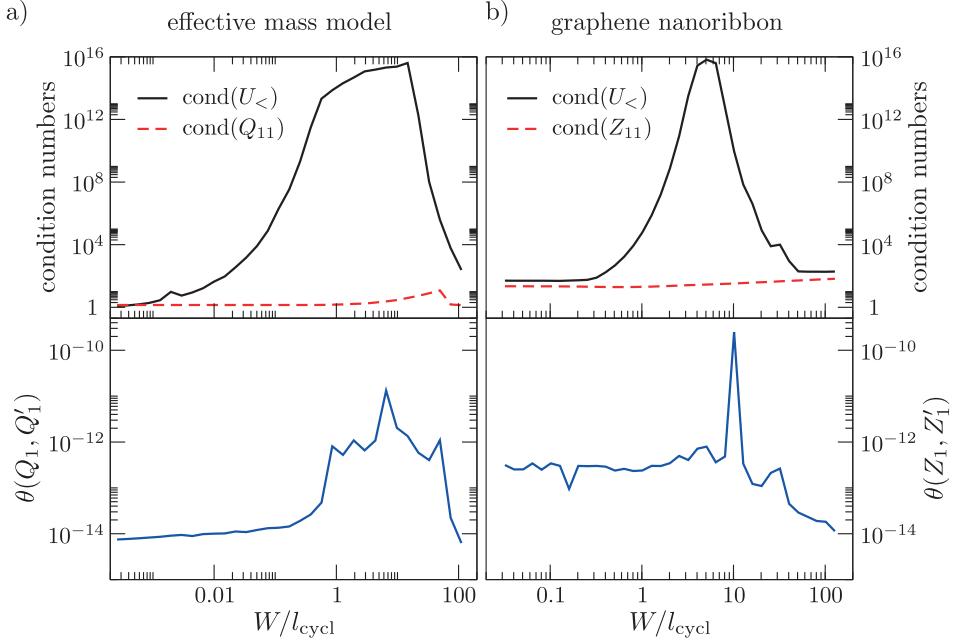


Figure 3.6 – Condition numbers (upper panels) and error in the computed invariant subspaces (lower panels) for the eigendecomposition and Schur based algorithms for the computation of the lead Green’s function. The quantities are shown as a function of W/l_{cycl} for two benchmark systems involving a magnetic field B , parameters as in Fig. 3.4. (a) a wire in effective mass approximation in a perpendicular magnetic field B , parameters as in Fig. 3.4. (b) a zigzag graphene nanoribbon in a perpendicular magnetic field B , with a width W corresponding to 302 atoms across the nanoribbon, and energy $E = 0.39t$.

3.3.4 Examples

In order to analyze their numerical performances, we apply both the eigendecomposition and Schur decomposition based algorithm to different systems. As a benchmark for the algorithms for invertible hopping matrices we take up again the system of Section 3.3.1, a wire in a magnetic field. For the case of singular hopping matrices, we consider a zigzag graphene ribbon described in nearest-neighbor tight-binding approximation⁹ in a perpendicular, homogeneous magnetic field B . In both cases we will again measure the magnetic field strength as W/l_{cycl} .

In the upper panels of Fig. 3.6 we show the condition numbers for the matrices that need to be inverted in the various algorithms: $U_<$ in the eigendecomposition

⁹ The tight-binding Hamiltonian is given by $H = \sum_{ij} t |i\rangle \langle j|$, where the summation runs over nearest neighbors only, and $t \approx 2.7\text{eV}$. For details on the model for graphene, see Appendix G

based algorithms, and Q_{11} , Z_{11} in the Schur decomposition based algorithms. In both examples, effective mass approximation and graphene, $\text{cond}(U_<) \approx 10^{10} - 10^{16}$ in a rather large range of magnetic field strength. In contrast, the matrices Q_{11} and Z_{11} are always well-behaved, with condition numbers smaller than 10^2 . As a consequence, the inversion of Q_{11} and Z_{11} is not problematic numerically, whereas the corresponding inversion of $U_<$ fails in the eigendecomposition based algorithms.

In the previous section we argued that the accuracy of the computed invariant subspace in the Schur decomposition based algorithms is expected to be superior to the eigendecomposition approaches. In order to demonstrate this, we define the angle between subspaces as a measure of the error in the computed subspace:

$$\theta(V, W) = \max_{\substack{v \in V \\ v \neq 0}} \max_{\substack{w \in W \\ w \neq 0}} \theta(v, w) \quad \text{for subspaces } V, W \text{ with } \dim V = \dim W. \quad (3.45)$$

In the lower panels of Fig. 3.6 we show the angle of the subspaces spanned by the columns of the computed Q_{11} (Z_{11}) and the true Q'_{11} (Z'_{11}). As expected, the accuracy of the subspaces computed from the Schur decomposition is very high, in contrast to the eigendecomposition that even yielded subspaces with the wrong dimensionality. Consequently, the Schur decomposition based algorithms perform very well for the whole magnetic field range.

Finally, we discuss the run times of the different algorithms. Note that the first step of computing numerically the eigendecomposition also involves computing the Schur decompositions [107]. This computational step is most time-consuming in both algorithms. As a consequence, we find similar run times: sometimes the Schur decomposition based algorithm is faster, sometimes the eigendecomposition based algorithm, depending on the system under consideration. However, we found only differences up to 20%, and thus both algorithms can be considered equivalent with respect to run time.

3.4 Summary

In this chapter we have derived expressions for calculating the surface Green's function of a lead valid for arbitrary tight-binding models, including singular hopping matrices. A direct implementation of these eigendecomposition based expressions has been found numerically unstable for certain cases, and an alternative algorithm based on the Schur decomposition has been developed to remedy these problems. Since the Schur decomposition based algorithm is in general more stable than the eigendecomposition based algorithm and of comparable speed, it should be the method of choice.

In the literature, there is only one other algorithm that can compute the lead Green's function for arbitrary tight-binding models [86].¹⁰ There, the problem of computing

¹⁰In fact, when the work on the contents of this chapter was started, there was no general algorithm available. Ref. [86] was published after the work on the algorithms of this chapter was finished.

the surface Green's function for a lead with singular hopping matrices is mapped to an equivalent problem with invertible hopping matrices. The algorithms of this chapter are conceptually much simpler and require fewer computational steps than the algorithm of Ref. [86]. Still, it might be interesting to compare the accuracy of these algorithms on different models. In any case, the algorithm of Ref. [86] can also benefit from the stability of the Schur decomposition based algorithm developed in this chapter.

For transport properties, it is necessary to calculate Green's functions at real energies E , and thus the derivations of this chapter assumed E to be real. For computing equilibrium densities, it is however often convenient numerically to compute Green's functions energies E in the complex plane [84, 113]. When E is complex there are no propagating modes, and the eigenmodes \mathbf{u}_n can be grouped into equally large sets of left- and right-decaying modes [97]. From these the lead Green's functions can be calculated as described in this chapter.

Chapter 4

Optimal block-tridiagonalization of matrices for quantum transport

4.1 Introduction

From the lead self-energies Σ_l^r derived in the previous chapter, we can now turn to calculating the retarded Green's function G_S^r of the scattering region. As discussed at the end of Chapter 2, the direct calculation of G_S^r by means of Eq. (2.111).

$$G_S^r = (E - H_S - \Sigma_l^r)^{-1}, \quad (4.1)$$

is usually computationally intractable. There is however a number of quantum transport algorithms for two-terminal systems—i.e. systems with only two leads, that we

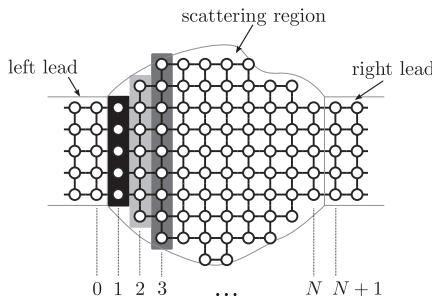


Figure 4.1 – Natural ordering of lattice points in a square finite difference grid leading to a block-tridiagonal matrix structure. The different matrix blocks are marked in alternating shades of grey.

denote as left (L) and right (R)—that circumvent this problem by making use of the sparsity of the Hamiltonian matrix H_S . In particular, these algorithm are based on the fact that the full Hamiltonian matrix H can be written in block-tridiagonal form:

$$H = \begin{pmatrix} & & & & \\ & \ddots & & & \\ H_L & V_L & & & \\ V_L^\dagger & H_L & H_{0,1} & & \\ & & & \ddots & \\ H_{1,0} & H_{1,1} & H_{1,2} & & 0 \\ & H_{2,1} & H_{2,2} & H_{2,3} & \\ & & H_{3,2} & \ddots & \\ & & & \ddots & H_{N-1,N} \\ & 0 & & H_{N,N-1} & H_{N,N} & H_{N,N+1} \\ & & & & H_{N+1,N} & H_R & V_R \\ & & & & V_R^\dagger & H_R & \\ & & & & & & \ddots \end{pmatrix}, \quad (4.2)$$

where the index L (R) denotes the blocks in the left (right) lead, 1...N the blocks within the scattering region S, and 0 ($N+1$) the first block in the left (right) lead.

This block-tridiagonal form of the Hamiltonian for two-terminal systems is the foundation of several quantum transport algorithms: For example, the transfer matrix approach applies naturally to block-tridiagonal Hamiltonians, but becomes unstable for larger systems. However, a stabilized version has been developed by Usuki *et al.* [92, 114]. In the decimation technique [115, 116], the Hamiltonian of the scattering region is replaced by an effective Hamiltonian between the two leads by eliminating internal degrees of freedom block by block. The contact block reduction method [117] calculates the full Green's function of the system using a limited set of eigenstates. The recursive Green's function (RGF) technique [103, 118, 119] uses Dyson's equation to build up the system's Green's function block by block. It has also been adapted to Hall geometries with four terminals [120] and to calculate non-equilibrium densities [121, 122]. Furthermore, the RGF algorithm has been formulated in a way suitable for parallel computing [123].

The block-tridiagonal form of H arises naturally, for example, in the method of finite differences, when grid points are grouped into vertical slices according to their x -coordinates, as shown in Fig. 4.1. In fact most of the above mentioned quantum transport algorithms are based on this *natural ordering*. Because of this, these algorithms are in general restricted to systems with two collinear leads. In a more complicated system, e.g. with non-collinear or multiple leads, a block-tridiagonal form is not obvious or involves only a few, very large blocks. In these cases, a direct application of the above mentioned transport algorithms is either very inefficient or even impossible.

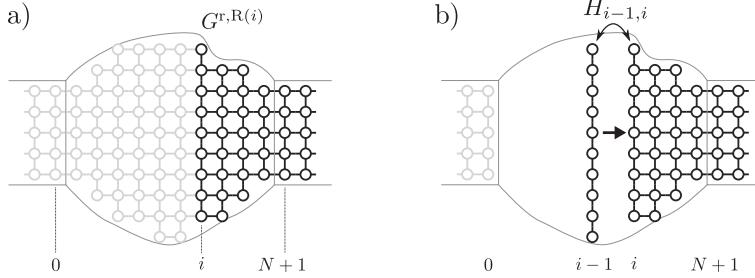


Figure 4.2 – Schematic depiction of the recursive Green’s function algorithm: (a) The Green’s function $G^{r,R(i)}$ contains all blocks $\geq i$. (b) The Green’s function $G^{r,R(i-1)}$ is obtained by adding another matrix block.

The intent of this chapter is to lift these restriction.

Of course, there are also other transport techniques not directly based on the block-tridiagonal form of the Hamiltonian matrix, such as extracting the Green’s function from wave packet dynamics [124]. Still, such algorithms are not as widely used as the large class of algorithms, that are directly based on the block-tridiagonal form of the Hamiltonian. In order to illustrate the typical computational tasks of this class of algorithms, we briefly state, as a representative example, the computational steps of the RGF algorithm. The details of this algorithm, including its extension to non-equilibrium, can be found in Appendix B.

The RGF technique is based on the Dyson’s equation $G^r = G_0^r + G_0^r V G^r$, building up the system block by block, as depicted in Fig. 4.2. Let $G^{r,R(i)}$ denote the Green’s function for the system containing all blocks $\geq i$. Then, the Green’s function $G^{r,R(i-1)}$ is related to $G^{r,R(i)}$ by

$$G_{i-1,i-1}^{r,R(i-1)} = \left(E - H_{i-1,i-1} - H_{i-1,i} G_{i,i}^{r,R(i)} H_{i,i-1} \right)^{-1} \quad (4.3)$$

and

$$G_{N+1,i-1}^{r,R(i-1)} = G_{N+1,i}^{r,R(i)} H_{i,i-1} G_{i-1,i-1}^{r,R(i-1)}. \quad (4.4)$$

Starting from $G_{N+1,N+1}^{r,R(N+1)} = g_R^r$, the surface Green’s function of the right lead, N slices are added recursively, until $G^{r,R(1)}$ has been calculated. The blocks of the Green’s function of the full system necessary for transport are then given by

$$G_{0,0}^r = \left((g_L^r)^{-1} - H_{0,1} G_{1,1}^{r,R(1)} H_{1,0} \right)^{-1} \quad (4.5)$$

and

$$G_{N+1,0}^r = G_{N+1,1}^{r,R(1)} H_{1,0} G_{0,0}^r, \quad (4.6)$$

where g_L^r is the surface Green’s function of the left lead. $G_{0,0}^r$ and $G_{N+1,0}^r$ are sufficient to calculate the conductance, as seen from Eqs. (2.117) and (C.31).

Each step of the algorithm performs inversions and matrix multiplications with matrices of size M_i . Since the computational complexity of matrix inversion and multiplications scales as M_i^3 , the complexity of the RGF algorithm is $\sum_{i=0}^{N+1} M_i^3$. Thus, it scales linearly with the “length” N , and cubically with the “width” M_i of the system. This scaling also applies to most of the other transport algorithms mentioned above.

While for particular cases general transport algorithms, such as the RGF algorithm, cannot compete with more specialized algorithms, such as the modular recursive Green’s function technique [125, 126] that is optimized for special geometries, they are very versatile and easily adapted to many geometries—provided that the system has only two leads that are arranged collinearly.

The objective of this chapter is twofold: First, we intend to lift the discussed restrictions of the established quantum transport algorithms. We do this by bringing the Hamiltonian matrix H into a block-tridiagonal form suitable for quantum transport also for complex geometries, such as non-collinear leads or multi-terminal structures, that would otherwise need the development of specialized algorithms. Second, we improve the matrix H such that the block-tridiagonal form is *optimal* for transport. Although the block-tridiagonal structure of H , Eq. (4.2), that arises naturally in many problems appears to have a small “width” and thus seems to be quite suitable for transport algorithms, optimizing the matrix structure further may lead to significant speed-ups even in the two-terminal case, as we show below.

We achieve these goals by developing a matrix reordering algorithm based on graph partitioning techniques, that brings an arbitrary tight-binding matrix H into a block-tridiagonal form optimal for quantum transport. To this end, we identify a mathematical measure for this optimality and reformulate the matrix reordering problem in the language of graph theory. Based on graph-theoretical concepts we then develop the block-tridiagonalization algorithm. Finally, we apply this algorithm to various examples and investigate performance of the reordering and of the RGF algorithm for the reordered Hamiltonian H , demonstrating the flexibility and power of this combined approach.

4.2 Definition of the problem

4.2.1 Definition of the matrix reordering problem

As shown in the previous section, the typical runtime of transport algorithms, proportional to $\sum_{i=0}^{N+1} M_i^3$, does depend on the particular block-tridiagonal structure of H . Therefore, the runtime of these algorithms can be improved in principle by conveniently reordering H with a permutation P ,

$$H' = P H P^{-1}. \quad (4.7)$$

In order to quantify the typical performance of a transport algorithm for a given

matrix structure, we define a weight $w(H)$ associated with a matrix H as

$$w(H) = \sum_{i=0}^{N+1} M_i^3, \quad (4.8)$$

where M_i is the size of block $H_{i,i}$. Optimizing the matrix for transport algorithms is then equivalent to minimizing the weight $w(H)$. Since $\sum_{i=0}^{N+1} M_i = N_{\text{grid}}$, where N_{grid} is the total number of grid points, $w(H)$ is minimal, if all M_i are equal, $M_i = N_{\text{grid}}/(N + 2)$. Therefore, a matrix tends to have small weight, if the number N of blocks is large, and all blocks are equally sized. The reordering problem of the matrix H is thus summarized as follows:

Problem 4.1. Matrix reordering problem: Find a reordered matrix H' such that

1. $H'_{0,0}$ and $H'_{N+1,N+1}$ are blocks given by the left and right leads (as required by transport algorithms)
2. H' is block-tridiagonal ($H'_{i,j} \neq 0$, iff $j = i + 1, i, i - 1$),
3. the number N of blocks is as large as possible, and all blocks are equally sized.

In principle, this constrained optimization problem could be solved by generic optimization algorithms, such as *Simulated Annealing*. However, for larger problems the optimization could take much more time than the actual transport calculation, rendering the optimization process useless. It is therefore necessary to use heuristics especially designed for the problem at hand. To this end, we formulate the matrix reordering problem in the language of graph theory.

4.2.2 Mapping onto a graph partitioning problem

A *graph* \mathcal{G} is an ordered pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is a set of *vertices* v and \mathcal{E} a set of ordered pairs of vertices $(v_1, v_2) \in \mathcal{V} \times \mathcal{V}$. Such a pair is called an *edge*. A graph is called *undirected*, if for every edge $(v_1, v_2) \in \mathcal{E}$ also $(v_2, v_1) \in \mathcal{E}$. Two vertices v_1 and v_2 are called *adjacent*, if $(v_1, v_2) \in \mathcal{E}$. In order to simplify the notation, we will also consider a vertex v to be adjacent to itself.

There is a natural one-to-one correspondence between graphs and the structure of sparse matrices. For a given $n \times n$ matrix H , we define a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{1, \dots, n\}$ and $(i, j) \in \mathcal{E}$ iff the entry $H_{ij} \neq 0$. A graph thus stores information about the *structure* of a matrix, i.e. which entries are nonzero. It does not contain any information about the values of the respective entries, although these may be stored easily along with the graph. However, for the formulation of the quantum transport algorithms, only the block-tridiagonal form, i.e. the structure of the matrix, is relevant. Hermitian matrices, that are considered in quantum transport, have a symmetric structure of zero and nonzero entries, and therefore the corresponding graphs are undirected.

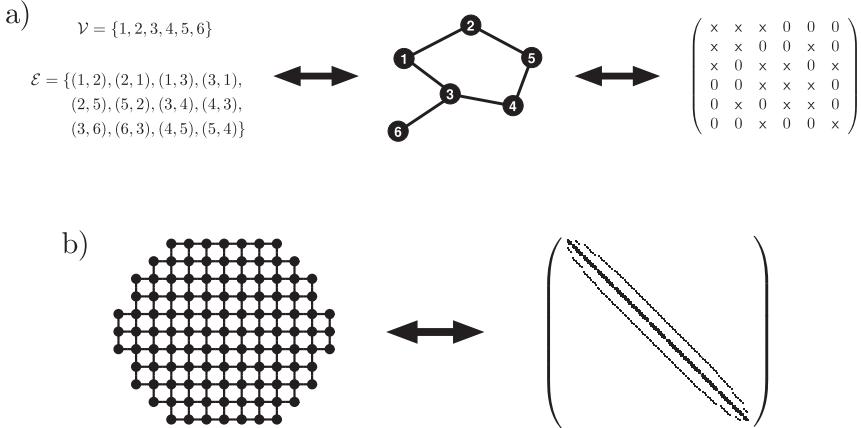


Figure 4.3 – (a) Simple example showing the connection between a graph, its graphical representations with dots and lines, and the zero–nonzero structure of a matrix. (b) Example of a finite difference grid, that can be interpreted as a graph, and the structure of the corresponding matrix. Nonzero entries are shown as black dots.

A graph can be depicted by drawing dots for each vertex v , and lines connecting these dots for every edge (v_1, v_2) , as shown in Fig. 4.3(a). It should be noted that a graphical representation of a tight-binding grid, such as shown in Fig. 4.3(b), can be directly interpreted as a representation of a graph and the corresponding matrix structure.

In terms of graph theory, matrix reordering corresponds to renumbering the vertices of a graph. Since we are only interested in reordering the matrix in terms of matrix blocks (the order within a block should not matter too much), we define a *partitioning* of \mathcal{G} as a set $\{\mathcal{V}_i\}$ of disjoint subsets $\mathcal{V}_i \subset \mathcal{V}$ such that $\bigcup_i \mathcal{V}_i = \mathcal{V}$ and $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for $i \neq j$. Using these concepts, we can now reformulate the original matrix reordering problem into a graph partitioning problem:

Problem 4.2. Graph partitioning problem: Find a partitioning $\{\mathcal{V}_0, \dots, \mathcal{V}_{N+1}\}$ of \mathcal{G} such that:

1. \mathcal{V}_0 and \mathcal{V}_{N+1} contain the vertices belonging to left and right leads,
2. (a) vertices in \mathcal{V}_0 and \mathcal{V}_{N+1} are only connected to vertices in \mathcal{V}_1 and \mathcal{V}_N , respectively,
 (b) for $0 < i < N + 1$, there are edges between \mathcal{V}_i and \mathcal{V}_j iff $j = i + 1, i, i - 1$,
3. the number $N + 2$ of sets \mathcal{V}_i is as large as possible, and all sets \mathcal{V}_i have the same cardinality $|\mathcal{V}_i|$. A partitioning with all $|\mathcal{V}_i|$ equally sized is called *balanced*.

A partitioning obeying requirement 4.2.2 is called a *level set* with *levels* \mathcal{V}_i [127]. Level sets appear commonly as an intermediate step in algorithms for bandwidth reduction of matrices [127–130]. These algorithms seek to find a level set of minimal

width, i. e. $\max_{i=0\dots N+1} |\mathcal{V}_i|$ as small as possible which is equivalent to requirement 4.2.3. The main difference between our graph partitioning problem and the bandwidth reduction problem is requirement 4.2.1: In the graph partitioning problem, \mathcal{V}_0 and \mathcal{V}_N are determined by the problem at hand, while in the bandwidth reduction problem these can be chosen freely. Due to this difference, bandwidth reduction algorithms can be applied successfully to our graph partitioning problem only for special cases, as we show below.

The term *graph partitioning* usually refers to the general problem of finding a balanced partitioning $\{\mathcal{V}_i\}$ of a graph and has many applications in various fields such as very-large-scale integration (VLSI) design [131–133], sparse matrix reorderings for LU or Cholesky decompositions [134], or block ordering of sparse matrices for parallel computation [135–140]. In particular, the latter examples also include block-tridiagonal orderings [136, 137]. However, as these reorderings are geared towards parallel computation, they obtain a fixed number N of sets \mathcal{V}_i given by the number of processors of a parallel computer, whereas in our block-tridiagonal reordering the number N should be as large as possible. In addition to that, the constraints on the blocks \mathcal{V}_0 and \mathcal{V}_{N+1} (requirement 4.2.1) are again not present there.

As we cannot directly employ existing techniques to solve the graph partitioning problem, we will develop an algorithm combining ideas from both bandwidth reduction and graph partitioning techniques in the subsequent sections: Concepts from bandwidth reduction are employed to construct a level set which is then balanced using concepts from graph partitioning.

4.3 Optimal Matrix reordering by graph partitioning

4.3.1 A local approach—breadth first search

A breadth-first-search (BFS) [141] on a graph immediately yields a level set [127–130]. In our particular example, the level set is constructed as follows:

Algorithm 4.1. Level set construction by breadth-first-search.

- A Start from $i = 0$. Then, $\mathcal{V}_i = \mathcal{V}_0$, as the first level is given by the constraints of requirement 4.2.1.
- B If there is a vertex in \mathcal{V}_i that is adjacent to a vertex in \mathcal{V}_{N+1} , assign all the remaining unassigned vertices into \mathcal{V}_{N+1} and end the algorithm.
- C All vertices adjacent to \mathcal{V}_i not contained in the previous levels $\mathcal{V}_i, \mathcal{V}_{i-1}, \dots, \mathcal{V}_0$ are assigned to \mathcal{V}_{i+1} .
- D Continue at step B with $i = i + 1$.

Note that the sets $\{\mathcal{V}_i\}$ form a level set by construction—a set \mathcal{V}_i may only have vertices adjacent to \mathcal{V}_{i-1} and \mathcal{V}_{i+1} . The construction by BFS not only obtains the

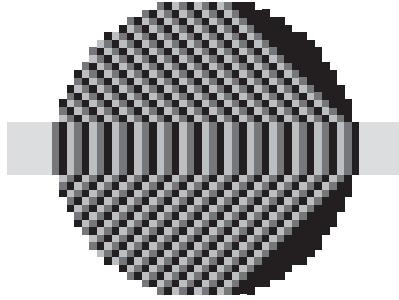


Figure 4.4 – Level set created by a BFS starting from \mathcal{V}_0 . Different levels are shown in alternating shades of grey.

number of levels $N+2$ for a particular realization, but yields a more general information:

Lemma 4.3. *The number of levels $N+2$ in the level set constructed by algorithm 4.1 is the maximum number of levels compatible with the constraints on the initial and final level \mathcal{V}_0 and \mathcal{V}_{N+1} for a graph \mathcal{G} .*

This can be seen from the fact that a BFS finds the shortest path in the graph between the initial sets \mathcal{V}_0 and \mathcal{V}_{N+1} , $(v_0, v_1, \dots, v_i, \dots, v_{N+1})$ where $v_0 \in \mathcal{V}_0$ and $v_{N+1} \in \mathcal{V}_{N+1}$. Any vertex on this shortest path can be uniquely assigned to a single level \mathcal{V}_i and it would not be compatible with a larger number of levels than $N+2$.

Algorithm 4.1 not only yields the maximum number of levels: All vertices contained in the first n levels of the BFS must be contained in the first n levels of *any* other level set:

Lemma 4.4. *Let $\{\mathcal{V}_0, \mathcal{V}_1, \dots, \mathcal{V}_{N+1}\}$ be a level set constructed by algorithm 4.1, and $\{\mathcal{V}'_0, \mathcal{V}'_1, \dots, \mathcal{V}'_{N'+1}\}$ another level set consistent with the requirements of problem 4.2 with $N' \leq N$. Then $\mathcal{V}_0 \cup \mathcal{V}_1 \cup \dots \cup \mathcal{V}_n \subset \mathcal{V}'_0 \cup \mathcal{V}'_1 \cup \dots \cup \mathcal{V}'_n$ for $0 \leq n \leq N'+1$.*

The statement is proved by induction. It is true trivially for $n = 0$ (because of requirement 1 in problem 4.2) and for $n = N'+1$ (then the levels cover the whole graph). Suppose now that the statement holds for $n < N'$. Note that for the proof it suffices to show that $\mathcal{V}_{n+1} \subset \mathcal{V}'_0 \cup \mathcal{V}'_1 \cup \dots \cup \mathcal{V}'_{n+1}$. Consider now the set of all vertices adjacent to \mathcal{V}_n , $\text{adjacent}(\mathcal{V}_n) = \{v \in \mathcal{V} \mid v \text{ is adjacent to a } v' \in \mathcal{V}_n\}$. By construction, $\mathcal{V}_{n+1} \subset \text{adjacent}(\mathcal{V}_n)$. Since $\mathcal{V}_n \subset \mathcal{V}'_0 \cup \mathcal{V}'_1 \cup \dots \cup \mathcal{V}'_n$ and $\{\mathcal{V}'_i\}$ is a level set, all vertices adjacent to \mathcal{V}_n must be contained in the set of vertices including the next level, i.e. $\text{adjacent}(\mathcal{V}_n) \subset \mathcal{V}'_0 \cup \mathcal{V}'_1 \cup \dots \cup \mathcal{V}'_n \cup \mathcal{V}'_{n+1}$. But then also $\mathcal{V}_{n+1} \subset \mathcal{V}'_0 \cup \mathcal{V}'_1 \cup \dots \cup \mathcal{V}'_{n+1}$, which concludes the proof.

Thus, the vertices contained in the first n levels of the BFS form a minimal set of vertices needed to construct n levels. However, this also implies that the last level which then covers the remaining vertices of the graph, may contain many more vertices

than the average, leading to an unbalanced level set. This is not surprising, since the algorithm does not explicitly consider balancing and only local information is used, i.e. whether a vertex is adjacent to a level or not. An example for this imbalance is shown in Fig. 4.4, where the BFS construction yields a very large last level.

Note that throughout the manuscript we visualize the graph theoretical concepts using examples of graphs obtained from discretizing a two-dimensional structure. However, the ideas and algorithms presented here apply to any graph and are not limited to graphs with coordinate information. Two-dimensional graphs have the advantage of being visualized easily. In particular, the BFS search has an intuitive physical analog: Wave front propagation of elementary waves emanating from the vertices of the initial level \mathcal{V}_0 .

The problem that a BFS does not yield a balanced partitioning was also noted in the theory of bandwidth reduction. The Gibbs-Poole-Stockmeyer (GPS) algorithm tries to overcome this deficiency by constructing a level set through the combination of two BFS searches starting from the initial and the final levels. However, there the initial and final levels are sought to be furthest apart, contrary to our problem. In general, the GPS construction only yields a balanced level set if the initial and final level are close to furthest apart, as we will show in Section 4.4.

4.3.2 A global approach—recursive bisection

In order to obtain a balanced partitioning, graph partitioning algorithms commonly perform a recursive bisection, i.e. successively bisect the graph and the resulting parts until the desired number of parts is obtained [131, 132, 136, 137, 142, 143]. This approach has the advantage of reducing the partitioning problem to a simpler one, namely bisection. Furthermore, if the resulting parts of every bisection are equally sized, the overall partitioning will be balanced. In addition, bisection is inherently a global approach, as the whole graph must be considered for splitting the system into two equally sized parts. Thus, it can be expected to yield better results than a local approach, such as BFS.

We intend to construct a level set with $N + 2$ levels, where $N + 2$ is the maximum number of levels as determined by algorithm 4.1. To this end we start from an initial partitioning $\{\mathcal{V}_0, \mathcal{V}_1, \mathcal{V}_{N+1}\}$, where \mathcal{V}_0 and \mathcal{V}_{N+1} contain the vertices of the leads (requirement 4.2.1), and \mathcal{V}_1 all other vertices. The level set is then obtained by applying the bisection algorithm recursively to \mathcal{V}_1 and the resulting subsets, until N levels are obtained, as shown schematically in Fig. 4.5. Here bisection means splitting a set \mathcal{V}_i into two sets, \mathcal{V}_{i_1} and \mathcal{V}_{i_2} , such that $\mathcal{V}_{i_1} \cup \mathcal{V}_{i_2} = \mathcal{V}_i$ and $\mathcal{V}_{i_1} \cap \mathcal{V}_{i_2} = \emptyset$. In order to be applicable to the graph partitioning problem 4.2, the bisection must comply with certain requirements:

Problem 4.5. The bisection algorithm must be

1. compatible with a level set with $N + 2$ levels.
2. balanced.

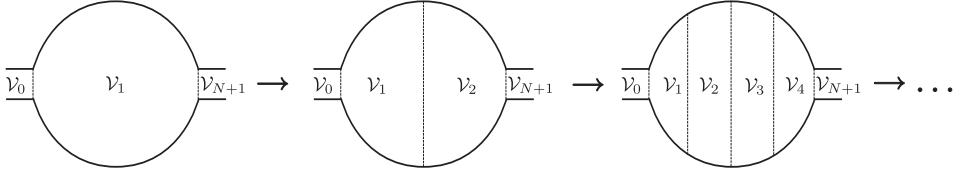


Figure 4.5 – Schematic depiction of recursive bisection.

3. performed such that subsequent bisections may lead to a balanced level set.

Requirement 4.5.3 is formulated rather vaguely: Usually there are many different choices how to perform a bisection. A particular choice will influence the subsequent bisections (for a similar problem in graph partitioning see [143]), and thus the bisection algorithm must in principle take into account all following bisection steps. Since an exact solution to that problem seems computationally intractable, we will resort to heuristics there.

We start explaining how to comply with requirements 4.5.1 and 4.5.2. In the following we assume that $N > 0$, as $N = -1, 0$ are trivial cases. Then the initial partitioning $\{\mathcal{V}_0, \mathcal{V}_1, \mathcal{V}_{N+1}\}$ forms a level set, and so will the final result of the recursive bisection, if the result of every intermediate bisection yields a level set. For this, consider a set \mathcal{V}_i with vertices adjacent to the sets $\mathcal{V}_{i_{\text{left}}}$ and $\mathcal{V}_{i_{\text{right}}}$, where “left” (“right”) is defined as being closer to \mathcal{V}_0 (\mathcal{V}_{N+1}). Then the sets resulting from the bisection, \mathcal{V}_{i_1} and \mathcal{V}_{i_2} may only have vertices adjacent to $\mathcal{V}_{i_{\text{left}}}, \mathcal{V}_{i_2}$ and $\mathcal{V}_{i_1}, \mathcal{V}_{i_{\text{right}}}$, respectively.

Apart from the condition of forming a level set, requirement 4.5.1 also dictates the total number of levels. Due to the nature of the recursive bisection, the number of final levels contained in an intermediate step is always well-defined. If a set \mathcal{V}_i contains N_i levels, then \mathcal{V}_{i_1} and \mathcal{V}_{i_2} must contain $N_{i_1} = \text{Int}(N_i/2)$ and $N_{i_2} = N_i - \text{Int}(N_i/2)$ levels, respectively. Here, $\text{Int}(\dots)$ denotes rounding off to the next smallest integer. The bisection is thus balanced, if

$$|\mathcal{V}_{i_1}| \approx \frac{N_{i_1}}{N_i} |\mathcal{V}_i| \quad \text{and} \quad |\mathcal{V}_{i_2}| \approx \frac{N_{i_2}}{N_i} |\mathcal{V}_i|. \quad (4.9)$$

Note that N_i can take any value, and usually is not a power of two.

From Lemma 4.4 we know that the minimum set of vertices necessary to form n levels is given by a BFS up to level n . Let $\mathcal{V}_{i_1,\text{BFS}}$ ($\mathcal{V}_{i_2,\text{BFS}}$) denote the set of vertices found by a BFS starting from $\mathcal{V}_{i_{\text{left}}}$ ($\mathcal{V}_{i_{\text{right}}}$) up to level N_{i_1} (N_{i_2}). Then, for any bisection complying with requirement 4.5.1, $\mathcal{V}_{i_1,\text{BFS}} \subset \mathcal{V}_{i_1}$ and $\mathcal{V}_{i_2,\text{BFS}} \subset \mathcal{V}_{i_2}$. These vertices are uniquely assigned to \mathcal{V}_{i_1} and \mathcal{V}_{i_2} and are consequently marked as *locked*, i.e. later operations may not change this assignment. An example for the vertices found in a BFS is shown in Fig. 4.6(a). Note that in the initial bisection, $\mathcal{V}_i = \mathcal{V}_1$, $N_i = N$, $\mathcal{V}_{i_{\text{left}}} = \mathcal{V}_0$, and $\mathcal{V}_{i_{\text{left}}} = \mathcal{V}_{N+1}$.

The remaining unassigned vertices can be assigned to either set, and the bisection will still be compatible with a level set containing $N + 2$ vertices. Thus for complying with

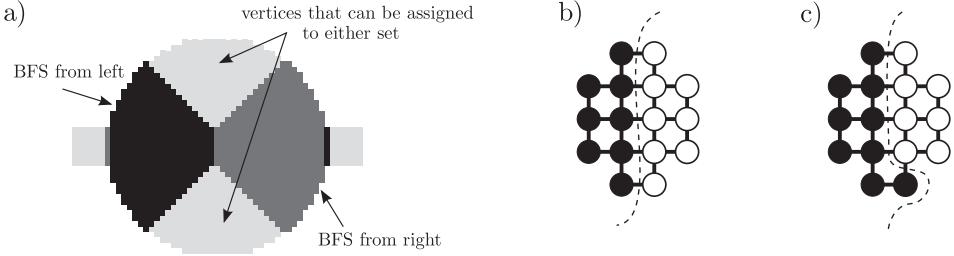


Figure 4.6 – (a) Example showing for a disk-type geometry the BFS from the left and right neighboring sets that construct the minimal set of vertices $\mathcal{V}_{i_1,\text{BFS}}$ (black) and $\mathcal{V}_{i_2,\text{BFS}}$ (dark grey) that must be contained in \mathcal{V}_{i_1} and \mathcal{V}_{i_2} , respectively. The remaining vertices (light grey) can be assigned to either set. (b) and (c): Examples illustrating the difference between cut edges and cut nets. The number of cut edges is 5 in both (b) and (c), while the number of cut nets (surface vertices) is 10 in (b) and 9 in (c).

requirement 4.5.2, any prescription obeying the balancing criterion may be used. We choose to distribute the remaining vertices by continuing the BFS from $\mathcal{V}_{i_{\text{left}}}$ and $\mathcal{V}_{i_{\text{right}}}$ and assigning vertices to \mathcal{V}_{i_1} and \mathcal{V}_{i_2} depending on their distance to the left or right neighboring set, while additionally obeying the balancing criterion. This approach—assigning vertices to levels according to their distance from the initial and final set—is rather intuitive and probably the procedure that would be used if the level set were to be constructed “by hand”. This procedure may lead to reasonable level sets, however in general, additional optimization on the sets \mathcal{V}_{i_1} and \mathcal{V}_{i_2} is needed, as discussed below. If this optimization is used, it can also be useful to distribute the unassigned vertices randomly, as this may help avoiding local minima.

As mentioned above, there is a lot of arbitrariness in distributing the unassigned vertices into \mathcal{V}_{i_1} and \mathcal{V}_{i_2} . However, the particular choice of the bisection will influence whether a later bisection is balanced or not: If $\mathcal{V}_{i_1(i_2),\text{BFS}}$ contains more vertices than given by the balance criterion (4.9), the bisection *cannot* be balanced. Obviously, the BFS that constructs $\mathcal{V}_{i_1(i_2),\text{BFS}}$ depends on the details of the set \mathcal{V}_i and thus on the details of the previous bisection step.

In order to formulate a criterion that may resolve the above-mentioned arbitrariness and help to find a balanced level set, it is useful to consider the matrix representation of the graph \mathcal{G} . Bisection a graph means ordering the corresponding matrix into two blocks that are connected by an off-diagonal matrix H_{i_1,i_2} :

$$\left(\begin{array}{c|c} \ddots & H_{i_1,i_2} \\ \hline H_{i_2,i_1} & \ddots \end{array} \right). \quad (4.10)$$

This off-diagonal matrix will be unchanged by further bisections and thus determines

the minimum level width that can be achieved. Therefore, the size of the off-diagonal matrix H_{i_1, i_2} should be minimized.

In a bisection, an edge $(v_1, v_2) \in \mathcal{E}$ is said to be *cut*, if v_1 and v_2 belong to different sets, i.e. $v_1 \in \mathcal{V}_{i_1}$ and $v_2 \in \mathcal{V}_{i_2}$ or vice versa. The entries of H_{i_1, i_2} correspond to edges cut by the bisection, and minimizing the number of entries in H_{i_1, i_2} corresponds to minimizing the number of edges cut by the bisection (*min-cut criterion*). This criterion is often used in reordering matrices for parallel processing, where the off-diagonal matrix size determines the amount of communication between processors.

However, the number of entries in H_{i_1, i_2} is not directly related to the size of the matrix, as has been noted in the graph partitioning problem for parallel computing [138]. Instead, the size of the off-diagonal matrix is given by the number of *surface* vertices, i.e. the number of vertices that have cut edges. For this, we define a *net* of a vertex v in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ as [136, 137]

$$\text{net}(v) = \{u \in \mathcal{V} | u \text{ is adjacent to } v\}. \quad (4.11)$$

Note that $v \in \text{net}(v)$, as v is adjacent to itself. A net is said to be cut by a bisection, if any two vertices $v_1, v_2 \in \text{net}(v)$ are contained in different sets \mathcal{V}_{i_1} and \mathcal{V}_{i_2} . Then, the number of surface vertices and thus the size of the off-diagonal matrix H_{i_1, i_2} is given by the number of cut nets. Thus, minimizing the number of cut nets (*min-net-cut criterion*) corresponds to minimizing the the number of surface vertices, and thus to minimizing the size of the off-diagonal matrix H_{i_1, i_2} . Furthermore, since the vertices in $\mathcal{V}_{i_{1/2}, \text{BFS}}$ are determined by a BFS emanating from the surface vertices, minimizing the number of cut nets will usually also lead to a smaller number of vertices in $\mathcal{V}_{i_{1/2}, \text{BFS}}$, leaving more freedom towards achieving a balanced bisection. Figs. 4.6(b) and (c) show a comparison of the min-cut and min-net-cut criterion for simple examples. In practice, when minimizing the number of cut nets, we also use the min-cut criterion to break ties between different bisections with the same number of cut nets (*min-net-cut-min-cut criterion*) in order to avoid wide local minima, that occur frequently in the min-net-cut problem.

Both the min-cut and min-net-cut bisection problem have been shown to be \mathcal{NP} -hard [144]. Therefore, only heuristics are available to solve them. These heuristics start from an initial (balanced) bisection, such as constructed by the steps outlined above, and improve upon this initial bisection. Here, we choose to use the Fiduccia-Mattheyses (FM) algorithm [132], as it is readily available for min-cut and min-net-cut bisection¹. Furthermore, the FM algorithm can naturally deal with locked vertices that may not be moved between sets, is reasonable fast and its underlying concepts are easy to understand. The FM heuristic is a *pass-based* technique, i.e. it is applied repeatedly to the problem (several *passes* are performed), iteratively improving the bisection. More detailed information about the fundamentals of the Fiduccia-Mattheyses algorithm are given in appendix E.

¹ In fact, min-net-cut bisection is a *hypergraph* partitioning problem, and the FM algorithm was originally designed for this task. For more information on hypergraphs, see Appendix E.

We now summarize the steps outlined above and formulate an algorithm for bisection:

Algorithm 4.2. Bisection of set \mathcal{V}_i containing N_i levels, with left (right) neighboring set $\mathcal{V}_{i_{\text{left}}}$ ($\mathcal{V}_{i_{\text{right}}}$).

- A Stop, if $N_i = 1$.
- B Do a BFS starting from $\mathcal{V}_{i_{\text{left}}}$ up to level $N_{i_1} = \text{Int}(N_i/2)$ and a BFS starting from $\mathcal{V}_{i_{\text{right}}}$ up to level $N_{i_2} = N - \text{Int}(N_i/2)$. The vertices found by the BFS are assigned to \mathcal{V}_{i_1} and \mathcal{V}_{i_2} , respectively, and are marked as locked.
- C Distribute the remaining unassigned vertices taking into account the balance criterion (4.9). The vertices may be assigned according to either one of the following prescriptions:
 - a) Continue the BFSs from step B and assign vertices to \mathcal{V}_{i_1} , if they are first reached by the BFS from $\mathcal{V}_{i_{\text{left}}}$, and to \mathcal{V}_{i_2} , if they are first reached by the BFS from $\mathcal{V}_{i_{\text{right}}}$. If a set has reached the size given by the balance criterion, assign all remaining vertices to the other set.
 - b) Distribute the unassigned vertices randomly to \mathcal{V}_{i_1} and \mathcal{V}_{i_2} . If a set has reached the size given by the balance criterion, assign all remaining vertices to the other set.
- D Optimize the sets \mathcal{V}_{i_1} and \mathcal{V}_{i_2} by changing the assignment of unlocked vertices according to some minimization criterion. In particular, the following optimizations may be performed:
 - a) No optimization.
 - b) Min-cut optimization using the FM algorithm.
 - c) Min-net-cut optimization using the FM algorithm.
 - d) Min-net-cut-min-cut optimization using the FM algorithm.

Recursive application of the bisection algorithm 4.2 then leads to an algorithm for constructing a level set complying with the requirements of the graph partitioning problem 4.2, and thus an algorithm for block-tridiagonalizing a matrix.

Algorithm 4.3. Block-tridiagonalization of matrix H

- A Construct the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ corresponding to the matrix H , and the sets \mathcal{V}_0 and \mathcal{V}_{N+1} corresponding to the leads.
- B Use algorithm 4.1 to determine the maximum number of levels $N + 2$. If $N < 1$, stop.
- C Construct $\mathcal{V}_1 = \mathcal{V} \setminus (\mathcal{V}_0 \cup \mathcal{V}_{N+1})$, containing N levels.
- D Apply the bisection algorithm 4.2 to \mathcal{V}_1 and then recursively on the resulting subsets. Do not further apply if a set only contains one level.

It should be emphasized, that the block-tridiagonalization does not require any other input than the graph structure. In principle, the number of FM passes may affect the result. However, from experience, this number can be chosen as a fixed value, e.g. 10

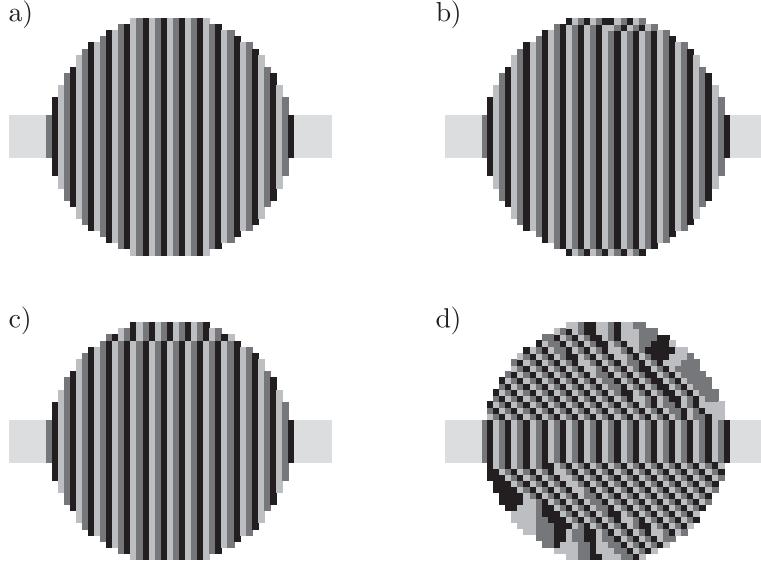


Figure 4.7 – Examples of level sets arising from (a) the natural ordering of grid points (as in Fig. 4.1), and application of the block-tridiagonalization algorithm 4.3 with distribution of vertices by BFS (algorithm 4.2, step C.(a)) (b) without further optimization, (c) with min-cut optimization, (d) with min-net-cut optimization.

FM passes, for all situations [132]. Thus, the block-tridiagonalization algorithm can serve as a black box.

In Fig. 4.7 we show examples of level sets arising from the natural ordering of grid points (Fig. 4.7(a), *natural level set*) and from the block-tridiagonalization algorithm developed in this work (Fig. 4.7(b)–(d)) for the case of a disk-type geometry. The level set in Fig. 4.7(b) arises from recursive bisection, where the vertices were distributed according to a BFS without any optimization. The resulting level set strongly resembles the natural level set. This is due to the highly symmetric structure and the fact that vertices are assigned to levels according to their distance from the leads—only small deviations are present due to the balance criterion. When the bisection is optimized according to the min-cut criterion, Fig. 4.7(c), the resulting level set changes little, as the min-cut criterion favors horizontal and vertical cuts for a square lattice, as presented in the example. In contrast, min-net-cut optimization (Fig. 4.7(d)) yields a new, non-trivial level set that has less symmetry than the underlying structure. Note that the minimization of surface vertices leads to levels in the form of ‘‘droplets’’, analogous to surface tension in liquids.

In fact, we will show in Section 4.4 that min-net-cut optimization usually leads to level sets and thus block-tridiagonal orderings that are superior to those arising from

other methods. In particular, they are better than the natural level sets, leading to a significant speed-up of transport algorithms, as demonstrated in Section 4.4.1. In addition to that, the reordering algorithms allow one to use conventional two-terminal transport algorithms also for more complicated, even multi-terminal structures (see Sections 4.4.1 and 4.4.2).

4.3.3 Computational complexity

We conclude the theoretical considerations with an analysis of the computational complexity of algorithms 4.2 and 4.3.

The bisection algorithm involves a BFS search on \mathcal{V}_i , which scales linearly with the number of edges within \mathcal{V}_i , and thus has complexity $\mathcal{O}(|\mathcal{E}_i|)$, where \mathcal{E}_i is the set of edges within \mathcal{V}_i . In addition to that, a single optimization pass of the FM algorithm scales also as $\mathcal{O}(|\mathcal{E}_i|)$ [132]. Usually, a constant number of passes independent of the size of the graph is enough to obtain converged results, and therefore the optimization process of several FM passes is also considered to scale as $\mathcal{O}(|\mathcal{E}_i|)$. Thus, the full bisection algorithm also has complexity $\mathcal{O}(|\mathcal{E}_i|)$.

Usually, the number of edges per vertex is approximately homogeneous throughout the graph. Since the recursive bisection is a divide-and-conquer approach, the computational complexity of the full block-tridiagonalization algorithm is then $\mathcal{O}(|\mathcal{E}| \log |\mathcal{E}|)$ [141]. In typical graphs arising from physics problems, the number of edges per vertex is a constant, and the computational complexity can be written as $\mathcal{O}(N_{\text{grid}} \log N_{\text{grid}})$, where N_{grid} is the number of vertices in \mathcal{V} , or the size of the matrix H .

In contrast, many quantum transport algorithms, such as the recursive Green's function technique, scale as $\mathcal{O}(N(N_{\text{grid}}/N)^3) = \mathcal{O}(N_{\text{grid}}^3/N^2)$ in the optimal case of N equally sized matrix blocks (levels) of size N_{grid}/N . Often, the number of blocks (levels) $N \propto N_{\text{grid}}^\alpha$. Typically, to name a few examples, $\alpha = 1$ in one-dimensional chains, $\alpha = 1/2$ in two dimensions, and the transport calculation scales as $\mathcal{O}(N_{\text{grid}}^{3-2\alpha})$. Thus, except for the case of a linear chain, where $N = N_{\text{grid}}$ and matrix reordering is pointless anyway, the block-tridiagonalization algorithm always scales more favorably than the quantum transport algorithms. This scaling implies that the overhead of the matrix reordering in the transport calculation will become more negligible, the larger the system size.

4.4 Examples: Charge transport in two-dimensional systems

4.4.1 Ballistic transport in two-terminal devices

We now evaluate the performance of the block-tridiagonalization algorithm using representative examples from mesoscopic physics. The Schrödinger equation for the two-dimensional electron gas (2DEG) is usually transformed into a tight-binding problem

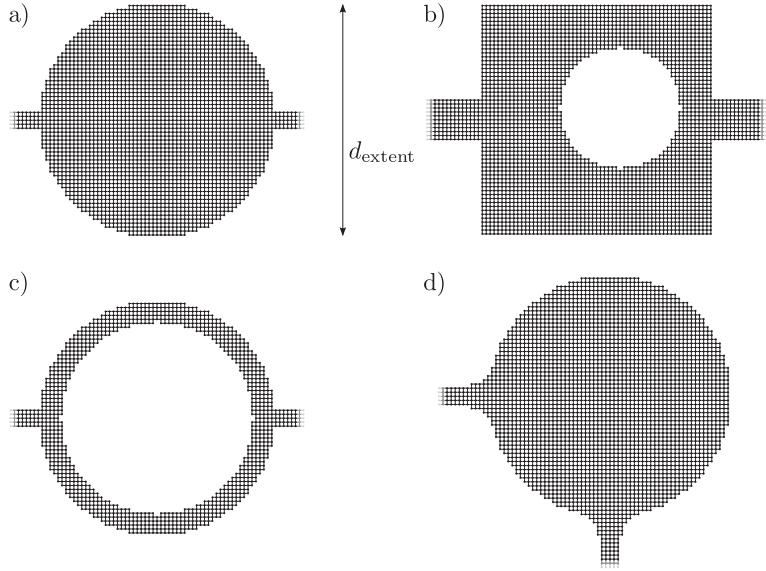


Figure 4.8 – Typical examples of structures considered in two-dimensional mesoscopic systems: (a) circle billiard, (b) asymmetric Sinai billiard, (c) ring, and (d) circular cavity with perpendicular leads. The tight-binding grid arises from the finite difference approximation to the Schrödinger equation. Note that the number of grid points used here was deliberately chosen very small for visualization purposes. In a real calculation, the number of grid points would be at least 2 orders of magnitude larger. d_{extent} denotes a length characterizing the extent for the different structures.

by the method of finite differences [77–79], where the continuous differential equation is replaced by a set of linear equations involving only the values of the wave function on discrete grid points, as discussed in Appendix F. Commonly, these points are arranged in a regular, square grid. This grid, together with the shape of the particular structure under consideration then defines the structure of the Hamilton matrix and the corresponding graph

The representative examples considered here are shown in Fig. 4.8: The circle (Fig. 4.8(a)) and the asymmetric Sinai billiard (Fig. 4.8(b)) that are examples of integrable and chaotic billiards in quantum chaos, respectively, the ring (Fig. 4.8(c)) that may exhibit various interference physics, and the circular cavity with leads that are not parallel (Fig. 4.8(d)) as an example of a structure that does not have an intuitive, natural block-tridiagonal ordering. For all these structures, we introduce a length scale d_{extent} , given by the outer radius of the circular structures and the side length of the square structure, characterizing the maximum extent. The fineness of the grid, and thus the size of the corresponding graph will be measured in number of grid points per

Table 4.1 – Weights $w(H)$, Eq. (4.8), for the block-tridiagonal ordering constructed by different algorithms for the examples of Fig. 4.8. Optimization was done by 10 passes of the FM algorithm, when the initial bisection was constructed by BFS (algorithm 4.2, step C.(a)), and 20 passes, when the initial bisection was constructed by a random distribution of vertices (algorithm 4.2, step C.(b)). The minimal weights for each system are printed bold. In all examples, there were 400 grid points per length d_{extent} .

	Circular billiard	Asymmetric Sinai billiard	Ring	Cavity with perp. leads
natural block-tridiagonal ordering	1.51×10^{10}	1.58×10^{10}	8.72×10^8	–
Gibbs-Poole-Stockmeyer	1.15×10^{12}	7.84×10^{11}	2.14×10^8	7.05×10^{12}
distribution by BFS, no optimization	1.51×10^{10}	9.29×10^9	2.1×10^8	1.69×10^{10}
distribution by BFS, min-cut	1.51×10^{10}	9.67×10^9	2.1×10^8	1.59×10^{10}
random distribution, min-cut	2.22×10^{10}	9.95×10^9	2.1×10^8	5.13×10^{10}
distribution by BFS, min-net-cut	1.51×10^{10}	9.46×10^9	2.1×10^8	1.18×10^{10}
random distribution, min-net-cut	1.46×10^{10}	9.0×10^9	2.09×10^8	1.18×10^{10}
distribution by BFS, min-net-cut-min-cut	1.26×10^{10}	9.28×10^9	2.08×10^8	1.24×10^{10}
random distribution, min-net-cut-min-cut	1.27×10^{10}	9.16×10^9	2.09×10^8	2.02×10^{10}

length d_{extent} .

We now apply the block-tridiagonalization algorithm using the various optimization criteria discussed in the previous section, and compare the resulting orderings with the natural ordering and the ordering generated by the GPS algorithm. The weights $w(H)$, Eq. (4.8), of the different orderings are given in Table 4.1.

The initial distributions for the bisection algorithm are done in two different ways: The vertices are distributed both in an ordered way—by BFS—and randomly. The outcome after the optimization however is always similar for both types of initial distributions which indicates that the resulting weights are close to the global minimum and not stuck in a local minimum. Note that we use twice as many FM passes for a random initial distribution than for an initial distribution by BFS, as convergence is usually slower for a random initial distribution.

In all examples, the min-net-cut criterion yields orderings with the best weights, as expected from the considerations of the previous section. Based on the weight, orderings according to this criterion are expected to give the best performance in transport calculations such as the RGF algorithm. Note that the min-net-cut-min-cut ordering is on average closest to the best ordering. The min-net-cut ordering sometimes

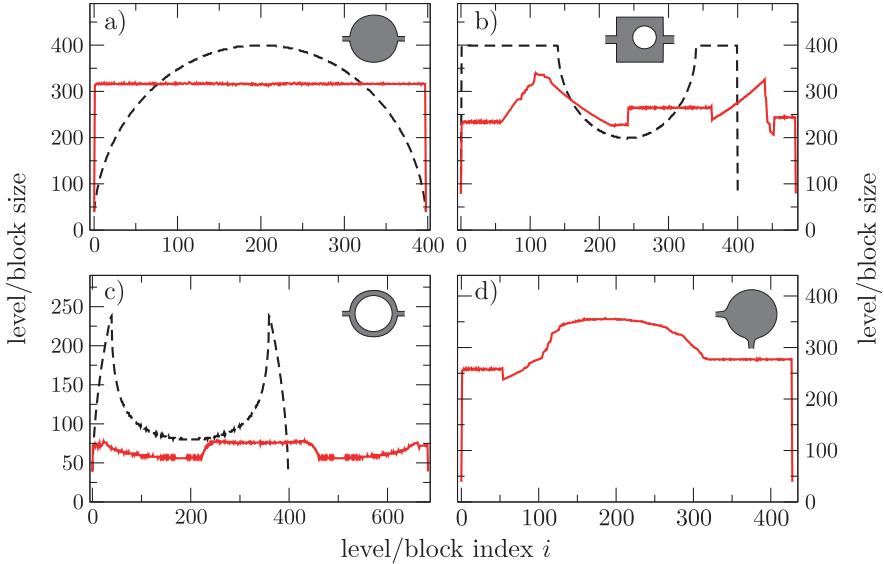


Figure 4.9 – Level (matrix block) size M_i as a function of the level (matrix block) index i for the natural level set (dashed line) and the min-net-cut-min-cut reordering (solid line), shown for (a) the circle billiard, (b) the asymmetric Sinai billiard, (c) the ring, and (d) the circular cavity with perpendicular leads. Note that for (d), there is no natural ordering. In all examples, there were 400 grid points per length d_{extent} .

suffers from slow convergence, when the algorithm must traverse a wide local minimum. The additional min-cut criterion helps to break ties and thus avoids these wide local minima.

Except for the ring, where all algorithms perform well, the GPS algorithm yields weights that are even larger than the weight of the natural ordering. As discussed above, the GPS algorithms performs well, if both leads are furthest apart in terms of the graph. In the case of the ring, this is approximately fulfilled. In the general case, when the leads are at arbitrary positions, the GPS algorithm usually produces some very large levels. As the level size enters cubically in the $w(H)$, this results in a prohibitively large weight. The GPS algorithm thus cannot be used as a generic reordering algorithm for quantum transport according to problem 4.2.

In summary, the block-tridiagonalization algorithm 4.3 in the combination of initial distribution by BFS and min-net-cut-min-cut optimization yields the best reorderings with respect to the weight $w(H)$. Experience shows that usually 10 FM passes are enough for optimizing a bisection. As a consequence, we will use this combination exclusively in the rest of this work.

The weight $w(H)$ of a matrix is a global measure of the quality of a ordering. Ad-

ditional insight can be gained from the distribution of the sizes M_i of the matrix blocks/levels. In Fig. 4.9 we show this distribution before and after reordering. For the natural ordering of the finite difference grids, the number of matrix blocks is determined by the number of lattice points along the x -coordinate direction (see Fig. 4.1(b)). In contrast, the number of matrix blocks after reordering is given by the length of the shortest path between the two leads, in terms of the corresponding graph.

In the case of the circle billiard, Fig. 4.9(a), the number of matrix blocks is the same for the natural ordering and the reordered matrix, as the shortest path between the leads is simply a straight line along the x -coordinate direction. The improvements in the weight originate only from balancing the matrix block sizes: While the matrix block sizes vary for the natural ordering—the lateral size changes along the x -direction—the reordered matrix has equally sized matrix blocks. For this particular example, the result of the block-tridiagonalization algorithm is optimal, as it yields the best solution with respect to the requirements set forth in problems 4.1 and 4.2. Note that it is not always possible to find a perfectly balanced partitioning, but the circle billiard is such an example.

In contrast, in the case of the asymmetric Sinai billiard and the ring the number of matrix blocks generated by the block-tridiagonalization algorithm is larger than in the natural ordering (see Figs. 4.9(b) and (c), respectively). In both cases, the obstacle within the scattering region increases the length of the shortest path connecting the two leads. In both examples, this increase in the number of matrix blocks leads to a significantly decreased weight $w(H)$ with respect to the natural ordering, although the partitioning is only approximately balanced. For instance, in the particular case of the ring, the number of matrix blocks after reordering is approximately given by the number of lattice points around half of the circumference. The reordered ring thus has a weight very similar to a straight wire with a width twice as large as the width of one arm of the ring, and a length given by half of the ring circumference.

For the cavity with perpendicular leads, there is no natural ordering, and a specialized transport algorithm would be required. The reordering creates a matrix with approximately balanced block sizes, and allows the direct application of conventional algorithms.

The weight $w(H)$ was introduced as a theoretical concept in order to simulate the computational complexity of a transport calculation. After discussing the influence of the reordering on this theoretical concept, we now demonstrate how the reordering increases the performance of an actual quantum transport calculation.

To this end we use a straight-forward implementation of the well-established recursive Green's function algorithm for two terminals, as described in Appendix B. It should be emphasized that the code that does the actual transport calculation—such as calculation of the Green's function and evaluation of the Fisher-Lee relation—is the same for all examples considered here, including the non-trivial cavity with perpendicular leads. The abstraction necessary for the reordering, i.e. the graph structure and the corresponding level set, allows for a generic computational code applicable to any tight-binding model.

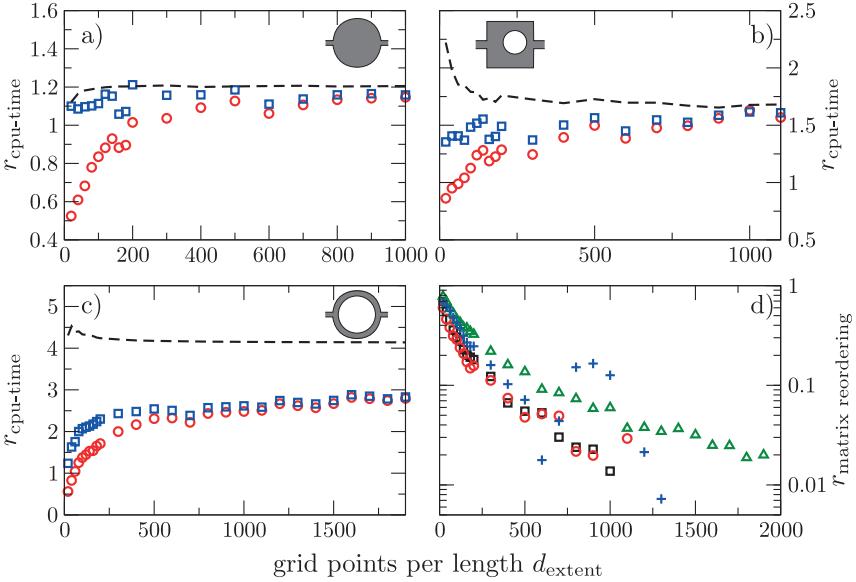


Figure 4.10 – (a)–(c): relative gain in computational time $r_{\text{cpu-time}}$, Eq. (4.12), through the reordering as a function of the grid size for the circular billiard, the asymmetric Sinai billiard, and the ring, respectively. $r_{\text{cpu-time}}$ is shown excluding (\square) and including (\circ) the overhead of matrix reordering. The estimate for $r_{\text{cpu-time}}$ from the weights $w(H)$ of the different orderings is shown as a dashed line. (d): fraction of time $r_{\text{matrix reordering}}$, Eq. (4.13), used for reordering the matrix as a function of the grid size. Data is shown for the circular billiard (\square), the asymmetric Sinai billiard (\circ), the ring (\triangle), and the circular cavity with perpendicular leads ($+$). The benchmarks were run on Pentium 4 processor with 2.8 GHz and 2 GBs of memory.

We measure the performance gain through matrix reordering as

$$r_{\text{cpu-time}} = \frac{\text{computing time for natural ordering}}{\text{computing time for reordered matrix}}. \quad (4.12)$$

Note that during a real calculation, the conductance is usually not only calculated once, but repeatedly as a function of some parameters, such as Fermi energy or magnetic field. Thus, the respective quantum transport algorithm is executed repeatedly, too. In contrast, the block-tridiagonalization has to be carried out again *only* when the structure of the matrix and thus the corresponding graph changes. For the examples considered here this would correspond to changing the grid spacing or the shape of the structure. In such a case, the overhead of matrix reordering must be taken into

account for $r_{\text{cpu-time}}$. This overhead can be quantified as

$$r_{\text{matrix reordering}} = \frac{\text{overhead of matrix reordering}}{\text{computing time including reordering}}. \quad (4.13)$$

In a typical calculation however, the matrix structure given by the underlying tight-binding grid does not change, and the matrix reordering must be carried out only once. In this common situation, the overhead of matrix reordering is negligible. For example, any change of physical parameters such as Fermi energy, magnetic field or disorder averages does not change the matrix structure.

In Fig. 4.10 we show the performance gain through matrix reordering, $r_{\text{cpu-time}}$, as a function of grid size for the circle billiard, the asymmetric Sinai billiard, and the ring (Figs. 4.10(a)–(c), respectively). We include both measurements excluding and including the overhead of matrix reordering, as discussed above. Remember that in the case of the cavity with perpendicular leads, Fig. 4.8(d), there is no natural ordering and thus a performance comparison is not possible. In fact for this system, only matrix reordering makes a transport calculation possible in the first place.

We find that block-tridiagonalization always increases the algorithmic performance in the typical situation, when the overhead of matrix reordering can be neglected. However, even if the reordering overhead is taken into account, we see a significant performance gain except for small systems—but there the total computing time is very short anyway. In fact, as the system sizes increases, the overhead of reordering becomes negligible, as predicted from the analysis of the computational complexity, and the performance gains including and excluding the reordering overhead converge. This can also be seen in Fig. 4.10(d), where we show the reordering overhead $r_{\text{matrix reordering}}$ as a function of system size.

Especially for large systems, the total computing time can become very long, and any performance gain is beneficial. Reordering leads to significant performance gains up to a factor of 3 in the case of the ring. The performance gain $r_{\text{cpu-time}}$ can also be estimated from the weights $w(H)$ of the original matrix (the natural ordering) and the reordered matrix, shown as the dashed line in Figs. 4.10(a)–(c). The actual, measured performance gain approaches this theoretical value, as the system size increases. Note that we do not fully reach the theoretically predicted performance gain in the case of the ring. On modern computer architectures, computing time does not only depend on the number of arithmetic operations [145], and thus the weight $w(H)$ overestimates the performance gain, though the performance still improves significantly.

Finally, we demonstrate the $\mathcal{O}(N_{\text{grid}} \log N_{\text{grid}})$ scaling of the reordering algorithm. Fig. 4.11 shows the computing times of the block-tridiagonalization algorithm as a function of matrix/graph size N for the geometries considered in this section. For all systems, the computing times scale according to the prediction from the complexity analysis in Section 4.3.3, as apparent from the fit $\propto N_{\text{grid}} \log N_{\text{grid}}$. Note that for large N_{grid} , $\mathcal{O}(N_{\text{grid}} \log N_{\text{grid}})$ scaling is practically indistinguishable from $\mathcal{O}(N_{\text{grid}})$ -scaling, as can also be seen in Fig. 4.11.

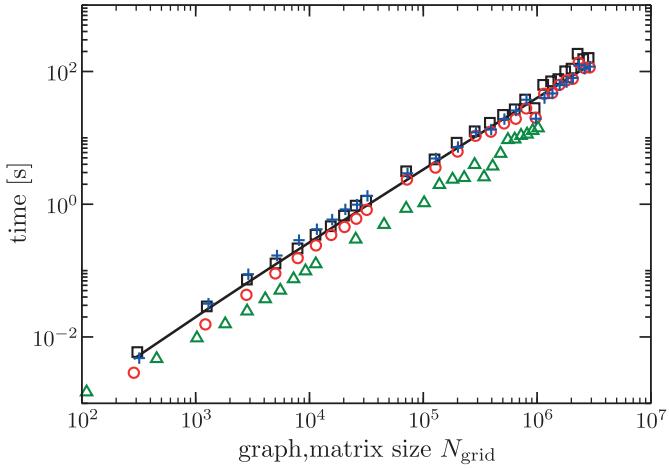


Figure 4.11 – Time spent for matrix reordering as a function of the total grid (matrix) size N , for the circular billiard (\square), the asymmetric Sinai billiard (\circ), the ring (\triangle), and the circular cavity with perpendicular leads ($+$). The solid line is a fit to the predicted scaling of the computational complexity, $N \log N$.

4.4.2 Multi-terminal structures

In the previous section, we demonstrated that matrix reordering increases the performance of quantum transport algorithms for two-terminal structures and additionally makes it possible to apply these conventional algorithms to non-trivial structures. Whereas there is a great variety of quantum transport algorithms for systems with two leads, there are only few algorithms that are suitable for multi-terminal structures, and most of them are restricted to rather specific geometries (e.g. Ref. [120]). Only recently algorithms have been developed that claim to be applicable to any multi-terminal structure. The *knitting algorithm* of Ref. [146] is a variant of the RGF algorithm where the system is built up adding every lattice point individually, instead of adding whole blocks of lattice points at a time. Therefore, instead of a matrix multiplication, the central computational step is an exterior product of vectors. Unfortunately, this implies that the knitting algorithm cannot use highly optimized matrix multiplication routines (Level 3 BLAS operations), that are usually much more efficient than their vector counterparts (Level 2 BLAS operations), as discussed in Ref. [145]. Another multi-terminal transport algorithm presented recently [147], is based on the transfer matrix approach. However, it requires the Hamiltonian to be in a specific block-tridiagonal form, and the corresponding level set is set up manually.

Here we show how to employ the block-tridiagonalization algorithm in order to apply the well-established two-terminal quantum transport algorithms to an arbitrary multi-

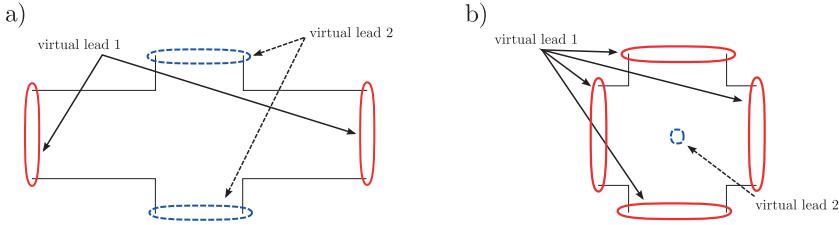


Figure 4.12 – A multi-terminal structure can be reduced to an equivalent two-terminal structure by collecting all leads in two *virtual leads*. (a) The leads are redistributed into two virtual leads. (b) All leads are combined in a single virtual lead, the second virtual lead is formed by a vertex furthest away.

terminal system. The basic idea is sketched in Fig. 4.12(a): Combining several (real) leads into only two virtual leads, the multi-terminal problem is reduced to an equivalent two-terminal problem. After reordering, the resulting problem can then be solved by conventional two-terminal algorithms. Note that in this approach the number of matrix blocks is given by the shortest path between leads in two different virtual leads. If all leads are very close together, this may lead to only few, large blocks in the reordered matrix and respectively levels in the graph partitioning, leading to a very large weight $w(H)$. In such a case it is advisable to collect all leads into a single virtual lead. The second virtual lead is then formed by a vertex in the graph, that is furthest away from all leads as depicted in Fig. 4.12(b). Such a vertex can be found by a BFS search originating from all leads. Thereby the number of matrix blocks/levels is maximized. In fact, this approach yields a block-tridiagonal matrix structure as required by the algorithm of Ref. [147].

We now demonstrate these strategies on the example of the quantum Hall effect (QHE) in a 2DEG formed in a semiconductor heterostructure [148] and in graphene [32]. For this we use a four-terminal Hall bar geometry as sketched in Fig. 4.12(a), on top of a square lattice (finite difference approximation to 2DEG, see Appendix F) and a hexagonal lattice (graphene, see Appendix G). Again, it should be emphasized that the code of the actual transport calculation is the same as employed in the two-terminal examples of the previous section. The results of the calculation are shown in Fig. 4.13, where the integer QHE of the 2DEG and the odd-integer QHE of graphene are clearly visible.

The methods outlined above make it possible to calculate quantum transport in *any* system described by a tight-binding Hamiltonian. This generality is one of the main advantages gained by using the matrix reordering. However, generality also implies that it is difficult to make use of properties of specific systems, such as symmetries, in order to speed up calculations. Special algorithms developed specifically for a certain system however can, and will usually be faster than a generic approach—at the cost of additional development time.

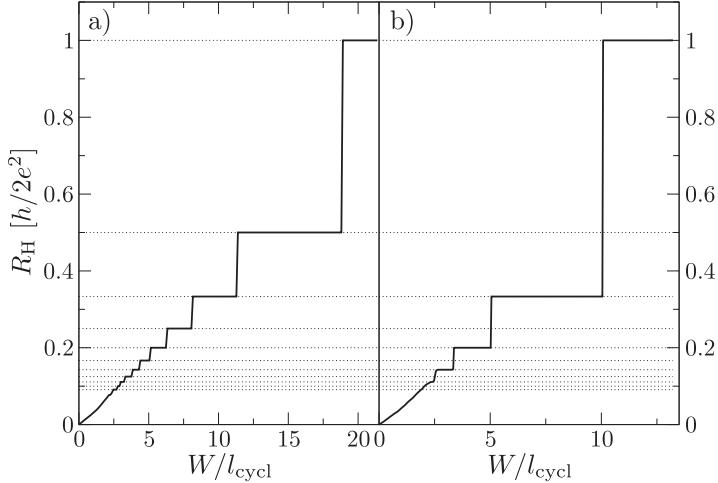


Figure 4.13 – Example of a four-terminal calculation: Quantum hall effect (a) in a two-dimensional electron gas and (b) in graphene. The Hall resistance R_H is shown as a function of W/l_{cycl} , where W is the width of the Hall bar and l_{cycl} the cyclotron radius in a magnetic field B . Note that $W/l_{\text{cycl}} \propto B$. The dotted lines indicate the quantized values of the Hall resistance, $h/2e^2 \times n^{-1}$, where n is a positive integer.

In the case of the Hall geometry in a 2DEG, such a special algorithm was presented by Baranger et al. [120], and we have implemented a variant of it. Comparing the computing times for the Hall bar geometry in a 2DEG, we find that the special algorithm is only a factor of 1.6 – 1.7 faster than our generic approach. Although such a performance comparison may depend crucially on the details of the system under consideration, experience shows that the use of the generic approach often does not come with a big performance penalty.

4.5 Summary

We have developed a block-tridiagonalization algorithm based on graph partitioning techniques that can serve as a preconditioning step for a wide class of quantum transport algorithms. The algorithm can be applied to any Hamiltonian matrix originating from an arbitrary tight-binding formulation and brings this matrix into a form that is more suitable for many two-terminal quantum transport algorithms, such as the widely used recursive Green's function algorithm. The advantages of this reordering are twofold: First, the reordering can speed up the transport calculation significantly. Second, it allows for applying conventional two-terminal algorithms to non-trivial geometries including non-collinear leads and multi-terminal systems. The

block-tridiagonalization algorithm scales as $\mathcal{O}(N_{\text{grid}} \log N_{\text{grid}})$, where N_{grid} is the size of the Hamiltonian matrix, and thus induces only little additional overhead. We have demonstrated the performance of the matrix reordering on representative examples, including transport in 2DEGs and graphene.

The framework set by the non-equilibrium Green's function formalism in tight-binding representation, presented in Chapter 2, provides a clear pathway for numerical transport calculations. Together with the novel results of the last two chapters, namely the general expression for the surface Green's function of the leads and the block-tridiagonalization algorithm, it is then possible to develop a transport code applicable to arbitrary tight-binding systems. This is possible, since both the calculation of the surface Green's function and the block-tridiagonalization algorithm can operate as a black box, i.e. without the need for specifying external parameters other than the tight-binding Hamiltonian H . Such a generic transport code is desirable, as it minimizes development time and increases code quality, as only a few basic, thoroughly tested transport routines are necessary.

The algorithms developed in this work have been implemented in a computer program, as a set of transport routines requiring only the input of the tight-binding Hamiltonian H . Adapting this code to an arbitrary tight-binding models thus only consists of appropriately setting up the matrix H . This has been done for the cases of a two- and three-dimensional effective mass model as well as for graphene, including the spin degree of freedom. In the remainder of this thesis we will employ this code to compute spin transport properties in magnetic tunnel junctions and graphene nanoribbons.

Part II

Spintronics in graphene and magnetic tunnel junctions

“The hype is bigger, because the physics is richer”

Carlo Beenakker – *The New York Times*, April 10, 2007

Chapter 5

Magnetic field effects on tunneling anisotropic magnetoresistance

5.1 Introduction

Amongst the various magnetoresistive phenomena, the giant magnetoresistance (GMR) [14, 15] and the tunneling magnetoresistance (TMR) effect [16–18] are probably the most prominent examples. In both cases, the resistance of two ferromagnetic layers separated by a spacer layer was found to depend on the relative angle of the magnetizations in the two ferromagnets. While this spacer is a metal in the GMR effect, it is an insulating material for the TMR effect. Although the basic experimental geometries are very similar in both cases, the underlying physics is quite different: diffusive transport in the case of GMR, and quantum mechanical tunneling in the case of TMR. Both the GMR and the TMR effect are employed in various electronic devices, such as read heads of hard disks or magnetic random access memory (MRAM) (for reviews see Refs. [12, 13, 19]).

In recent years, several experiments have found that the resistance of a magnetic tunnel junction (MTJ) may also depend on the *absolute* angle of the magnetization in a ferromagnet with respect to some crystallographic axis, in contrast to the conventional GMR and TMR effects, where the resistance depends only on the *relative* angle between the magnetizations in two ferromagnetic layers. Amongst those, the most striking are experiments on MTJs with only a single ferromagnetic layer that nevertheless show a magnetoresistance effect, depending on the orientation of the magnetization in the ferromagnet [47, 149]. This magnetoresistance effect is called *tunneling anisotropic magnetoresistance* (TAMR) effect.

The TAMR effect was first observed in a $(\text{Ga},\text{Mn})\text{As}/\text{Al}_2\text{O}_3/\text{Au}$ tunnel junction, featuring epitaxially grown magnetic semiconductor $(\text{Ga},\text{Mn})\text{As}$ as the ferromagnetic elec-

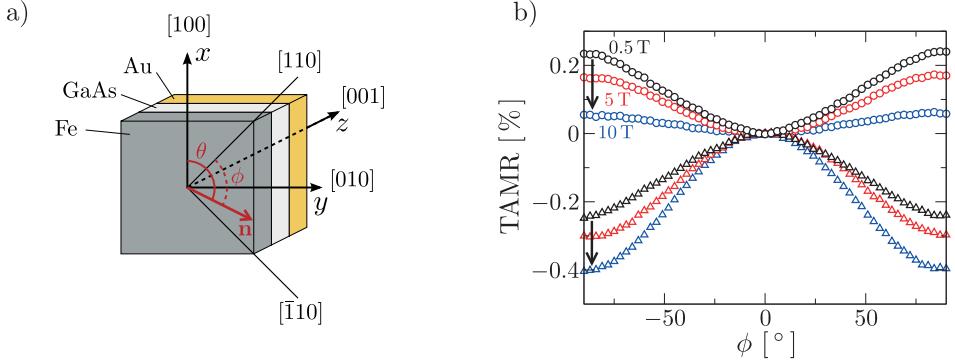


Figure 5.1 – (a) Schematic view of the magnetic tunnel junction in the experiments of Refs. [47, 150]. (b) TAMR effect for different bias voltages ($V_{\text{bias}} = +90 \text{ meV}$: \circ ; $V_{\text{bias}} = -90 \text{ meV}$: Δ) and magnetic fields (black: $B = 0.5 \text{ T}$; red: $B = 5 \text{ T}$; blue: $B = 10 \text{ T}$) in a Fe/GaAs/Au tunnel junction (from [150]).

trode, and an amorphous aluminum-oxide tunnel barrier [149]. The observed TAMR exhibited a four-fold anisotropy and was explained by an anisotropic density of states due to uniaxial strain in the magnetic semiconductor.

A qualitatively different TAMR effect was found in epitaxially grown Fe/GaAs/Au tunnel junctions with a single-crystalline tunnel barrier [47, 150], showing a uniaxial anisotropy in the magnetoresistance. The GaAs barrier was grown in the [001]-direction, as indicated in the schematic of the experiment in Fig. 5.1(a). The angular dependence of the TAMR effect can be quantified as

$$\text{TAMR}(\phi) = \frac{R(\phi) - R(0)}{R(0)}, \quad (5.1)$$

where $R(\phi)$ is the resistance of the tunnel junction, when the magnetization in Fe forms an angle ϕ with the [110]-direction, as indicated in Fig. 5.1(a). In the experiment, the TAMR was found to be periodic in ϕ , with a periodicity of π , as shown in Fig. 5.1(b). The TAMR effect shows both a peculiar bias as well as magnetic field dependence: For different bias voltages V_{bias} the TAMR changes in magnitude, and may also change sign. Moreover, the magnitude of the effect decreases with increasing magnetic field B when the TAMR effect is positive, and it increases when the TAMR effect is negative (magnetic field increases from the black to the blue symbols in Fig. 5.1(b)).

Theoretically, the TAMR effect for the case of small magnetic fields can be explained by the effects of spin-orbit interaction in the tunneling process [47, 151, 152]. This theory assumes that the momentum parallel to the layers (in-plane momentum) is conserved during the tunneling process, as appropriate for epitaxially grown (in contrast to amorphous) systems. The interference of Bychkov-Rashba spin-orbit coupling

at the Fe/GaAs interface and Dresselhaus spin-orbit coupling within the GaAs-barrier then leads to the observed uniaxial anisotropy of the magnetoresistance. The bias dependence of the TAMR is explained by a bias-dependent value of the interfacial Bychkov-Rashba spin-orbit parameter. However, up to now, a theoretical explanation of the magnetic field dependence of the TAMR is still missing. Such an explanation is the objective of this chapter.

To this end, the chapter is organized as follows: First, we extend the theoretical model of Refs. [47, 151, 152] to include the orbital effects of a magnetic field B . The resulting model is then solved numerically and compared to experimental data. Finally, we develop a qualitative model for the B -field dependence of the TAMR, reproducing all characteristic features of the experiment.

5.2 Magnetic field dependence of TAMR

5.2.1 Model

Within the theoretical model of the TAMR effect, the Hamiltonian reads [47, 151, 152]

$$H = H_0 + H_D + H_{\text{BR}}, \quad (5.2)$$

where

$$H_0 = \frac{1}{2} \mathbf{p} \frac{1}{m(z)} \mathbf{p} + V(z) - \frac{\Delta(z)}{2} \mathbf{n} \cdot \boldsymbol{\sigma}. \quad (5.3)$$

Here, $\mathbf{p} = -i\hbar\nabla$, and the coordinate system was chosen such that the z -direction is in the growth direction of the MTJ, as indicated in Fig. 5.1(a). $m(z)$ is the effective mass, with $m(z) = 0.067m_e$ inside the GaAs barrier, and $m(z) \approx m_e$ within the Fe and Au layer, where m_e is the bare electron mass. $V(z)$ is the conduction band profile and $\Delta(z)$ the spin splitting. We neglect the Zeeman splitting in the barrier and the non-magnetic electrode, as these are much smaller than any other energy scale in the system, and only keep the spin splitting due to the magnetization in the ferromagnet. The direction of this magnetization is given by

$$\mathbf{n} = \begin{pmatrix} \cos \theta \\ \sin \theta \\ 0 \end{pmatrix}, \quad (5.4)$$

where θ is the angle with respect to the x -axis. The experimentally used angle ϕ and θ are related through

$$\phi = \theta + \frac{\pi}{4}, \quad (5.5)$$

as seen from the sketch in Fig. 5.1(a). $\boldsymbol{\sigma}$ is the vector of Pauli matrices. The shape of the conduction band offset resulting from this model is shown schematically in Fig. 5.2(a). The barrier height is approximately given by half the GaAs band gap, so that $V_{\text{bar}} \approx 0.75$ eV. In addition, the Fermi wave vectors within Fe are $k_{\text{F,Fe}}^\uparrow = 1.05 \times 10^{10} \text{ m}^{-1}$ and

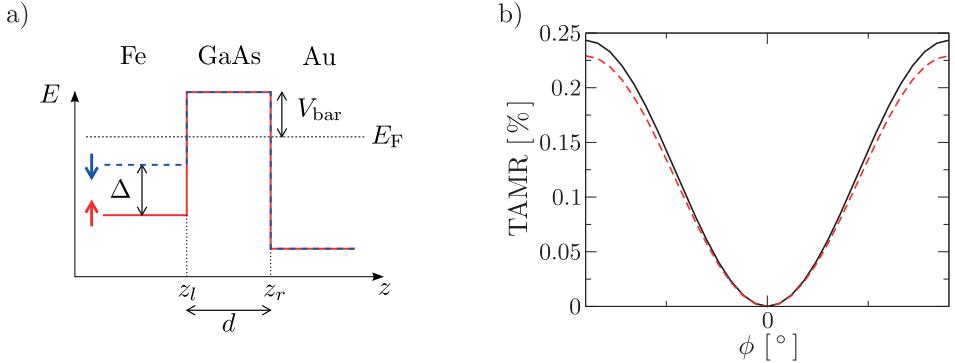


Figure 5.2 – (a) Schematic depiction of the conduction band profile for spin up (red solid line) and spin down (blue dashed line) for the theoretical model of the Fe/GaAs/Au tunnel junction. (b) Comparison between the TAMR calculated numerically from the finite difference approximation with $a_0 = 0.01 \text{ nm}$ (black solid line) and from an approximate analytical expression [152] (red dashed line). The Bychkov-Rashba parameter was chosen as $\alpha = 25 \text{ eV } \text{\AA}^2$.

$k_{\text{F},\text{Fe}}^\downarrow = 0.44 \times 10^{10} \text{ m}^{-1}$ for spin up and down, respectively. Inside the Au layer, the Fermi wave vector is given as $k_{\text{F},\text{Au}} = 1.2 \times 10^{10} \text{ m}^{-1}$.

H_{D} and H_{BR} describe the Dresselhaus and Bychkov-Rashba spin-orbit coupling, respectively. Here

$$H_{\text{D}} = \frac{1}{\hbar} (\sigma^x p_x - \sigma^y p_y) \frac{\partial}{\partial z} \left(\gamma(z) \frac{\partial}{\partial z} \right), \quad (5.6)$$

where the Dresselhaus parameter is $\gamma \approx 24 \text{ eV } \text{\AA}^3$ within the GaAs barrier and $\gamma = 0$ outside. The Bychkov-Rashba spin-orbit coupling is only relevant at the Fe/GaAs interface (which exhibits a broken inversion symmetry) and the Hamiltonian is given as

$$H_{\text{BR}} = \frac{1}{\hbar} \alpha (\sigma^x p_y - \sigma^y p_x) \delta(z - z_l), \quad (5.7)$$

where z_l is the position of the left interface. In contrast to the Dresselhaus parameter γ which is a bulk property, the interface Bychkov-Rashba parameter is not known. Instead, α is used as a fitting parameter to describe the TAMR at a given bias voltage V_{bias} , hence resulting in a bias-dependent Bychkov-Rashba parameter $\alpha(V_{\text{bias}})$. A summary of the material parameters used in the model is given in Table 5.1. Furthermore, in the remainder of this chapter we fix the barrier width $d = 8 \text{ nm}$, as in the experiments.

In order to include the orbital effects of a magnetic field \mathbf{B} into this model, we first

m_{GaAs}	m_{Fe}	m_{Au}	V_{bar}	$k_{\text{F},\text{Fe}}^{\dagger}$	$k_{\text{F},\text{Fe}}^{\dagger}$	$k_{\text{F},\text{Au}}$	γ_{GaAs}
$0.067 m_e$	m_e	m_e	0.75 eV	$1.05 \times 10^{10} \text{ m}^{-1}$	$0.44 \times 10^{10} \text{ m}^{-1}$	$1.2 \times 10^{10} \text{ m}^{-1}$	24 eV \AA^3

Table 5.1 – Summary of the material parameters of the theoretical model for the Fe/GaAs/Au tunnel junction (from Refs. [13, 47, 151]).

note that the magnetic field is antiparallel to \mathbf{n} , so that¹

$$\mathbf{B} = \begin{pmatrix} -B \cos \theta \\ -B \sin \theta \\ 0 \end{pmatrix}. \quad (5.8)$$

We then choose the vector potential such that

$$\mathbf{A}(z) = \begin{pmatrix} -B \sin \theta z \\ B \cos \theta z \\ 0 \end{pmatrix}, \quad (5.9)$$

where $\mathbf{B} = \nabla \times \mathbf{A}$. The magnetic field then enters the Hamiltonian via minimal coupling

$$\mathbf{p} \rightarrow \mathbf{p} + e\mathbf{A}(z). \quad (5.10)$$

This substitution must be done for any momentum operator in the Hamiltonian, including the spin-orbit terms.

In the chosen gauge, the vector potential in Eq. (5.9) only depends on the longitudinal coordinate z . Hence, the Hamiltonian is still translationally invariant in x and y -direction, and Bloch's theorem applies. Therefore, the full wave function can be written as

$$\varphi(\mathbf{x}) = e^{i(k_x x + k_y y)} \varphi(z), \quad (5.11)$$

and the Schrödinger equation for the Hamiltonian H reduces to a one-dimensional differential equation for $\varphi(z)$. The transverse wave vector $\mathbf{k}_{||} = (k_x, k_y)^t$ is a good quantum number, as in the case without magnetic field.

At this point it should be emphasized that the conservation of $\mathbf{k}_{||}$ during the tunneling process is *not* in contradiction to the cyclotron effects expected from a magnetic field. In fact, physically observable quantities, such as the electron velocity $\mathbf{v}_{||} = \frac{1}{m}(\hbar \mathbf{k}_{||} + e\mathbf{A}_{||}(z))$ do show the transverse acceleration expected from the Lorentz force².

¹ Magnetic moments $\boldsymbol{\mu}$ align *parallel* to the magnetic field \mathbf{B} , but the magnetic moment of an electron is *antiparallel* to its spin: $\boldsymbol{\mu}_e = -\frac{e}{2m_e} g \frac{\hbar}{2} \boldsymbol{\sigma}$, where $g \approx 2$ is the gyromagnetic factor of the electron.

² For further discussion on the difference between canonical and mechanical momentum, see, for example, Ref. [153, App. III]

The choice of gauge in Eq. (5.9) also comes with a disadvantage: It is not possible to include a magnetic field in the contacts³. In practice, we therefore switch off the magnetic field after some distance away from the barrier. This is a good approximation, as the effects of the magnetic field are expected to be strongest in the barrier and negligible in the metallic contacts: The cyclotron radius $l_{\text{cycl}} = p_F/eB$ [111] takes the values $l_{\text{cycl}} \approx 700$ nm in Fe and $l_{\text{cycl}} \approx 800$ nm in Au, much larger than any relevant length scale in the problem. Indeed, we find that the numerical results are insensitive to the exact position where the magnetic field is switched off.

In order to solve the Hamiltonian (5.2) including a magnetic field numerically, we employ the finite difference approximation that yields a tight-binding description of the system (for details, see Appendix F). The Delta function in Eq. (5.7) is approximated by a step function of width $2a_0$, centered at the Fe/GaAs interface. Here, a_0 is the lattice spacing of the discretization. In order to compare our numerical results with the previous analytical results of Refs. [47, 151, 152], we choose a very fine lattice spacing $a_0 = 0.01$ nm, to obtain a good approximation for the Delta function in Eq. (5.7).⁴

For a given $\mathbf{k}_{||}$, the transmission probability $T(E, \mathbf{k}_{||})$ at energy E can then be calculated using the numerical techniques developed in Part I. Then, the total conductance is given as

$$G(\phi) = \frac{S}{(2\pi)^2} \int d\mathbf{k}_{||} T(E, \mathbf{k}_{||}), \quad (5.12)$$

where S is the cross-section area of the magnetic tunnel junction. Note that the finite difference approximation as presented in Appendix F is not restricted to situations where $\mathbf{k}_{||}$ is conserved, but can be also applied to disordered systems [155]. In terms of conductances, the TAMR value is given as

$$\text{TAMR}(\phi) = \frac{R(\phi) - R(0)}{R(0)} = \frac{G(0) - G(\phi)}{G(\phi)}. \quad (5.13)$$

In Fig. 5.2(b) we show a comparison of the TAMR effect without magnetic field calculated both numerically and from an approximate analytical expression [152]. Given the fact that the analytical expression is only approximate, we find a very good agreement between the numerical and analytical results⁵ confirming the validity of the numerical procedure.

³ First of all, scattering states can only be defined if the vector potential is parallel to the direction of the lead [154]. Second, also the method to calculate the surface Green's function presented in Chapter 3 is based on translational invariance along the lead direction.

⁴ For the description of the experiment it is not necessary to choose such a fine lattice spacing. In fact, no length scales smaller than the inter-atomic distance should play a role. However, the value of the Bychkov-Rashba parameter α depends on the width $2a_0$ of the step function mimicking the Delta function of Eq. (5.7); hence a fine lattice spacing is needed to compare with the previous analytical results.

⁵ When comparing results from this thesis and previous work, note that Refs. [47, 151] used a different sign convention for α .

5.2.2 Numerical results

We now compare the results of numerical simulations on the magnetic field dependence of the TAMR effect with the corresponding experimental data. For this, we fit the parameter α at $B = 0.5$ T for every value of the bias voltage V_{bias} to the experimental data. In doing so, we obtain the bias voltage dependence of $\alpha = \alpha(V_{\text{bias}})$, as in Refs. [47, 151]. The magnetic field dependence can then be calculated without fitting any further parameter.

In Fig. 5.3(a) we show the angular dependence of the TAMR effect for different values of the bias voltage and magnetic field. The numerical simulations show the same trend as the experiment: The TAMR effect decreases with increasing B , when the effect is positive, and it increases, when the effect is negative. Furthermore, the numerical calculations reproduce the experimentally found change with magnetic field within a factor of 1.5 – 2. This is an especially satisfying agreement, given the fact that the magnetic field dependence is calculated without fitting any parameter.

In order to quantify the magnetic field dependence of the magnitude of the TAMR effect, we show $\text{TAMR}(90^\circ)$ as a function of the magnetic field B in Fig. 5.3(b). In the experiment, we find that $\text{TAMR}(90^\circ)$ changes linearly with B . Moreover, the slope of this change is approximately independent of the bias voltage V_{bias} . Both of these characteristic features are also found in the numerical simulations. As before, the numerical results underestimate the slope of the magnetic field dependence only by a factor of 1.5 – 2.

Up to now, detailed experimental data on the B -field dependence of $\text{TAMR}(90^\circ)$ is only available up to 5 T. From the magnetic field dependence of single TAMR curves as shown in Fig. 5.3(a), it seems that $\text{TAMR}(90^\circ)$ may change faster than linear in B for $B > 5$ T, although this is hard to decide from only a few points. In contrast, the numerical simulations result in a linear B -dependence also for fields up to 10 T. For large fields, other terms than those considered in Eq. (5.2) may become important [13]. In any case, the discussion of this point should be postponed until further experimental data is available.

The numerical results of this section show that the orbital effects of a magnetic field are enough to reproduce the characteristic features of the magnetic field dependence in the experiment. In order to understand the underlying physical mechanisms of this effect, we develop a qualitative model of the B -field dependence of the TAMR effect in the next section.

5.2.3 A qualitative picture

In order to develop a qualitative model for the magnetic field dependence of the TAMR, we extend the phenomenological model of Refs. [13, 151] to include finite magnetic fields.

The transmission probability can be written as a perturbative expansion in the spin-

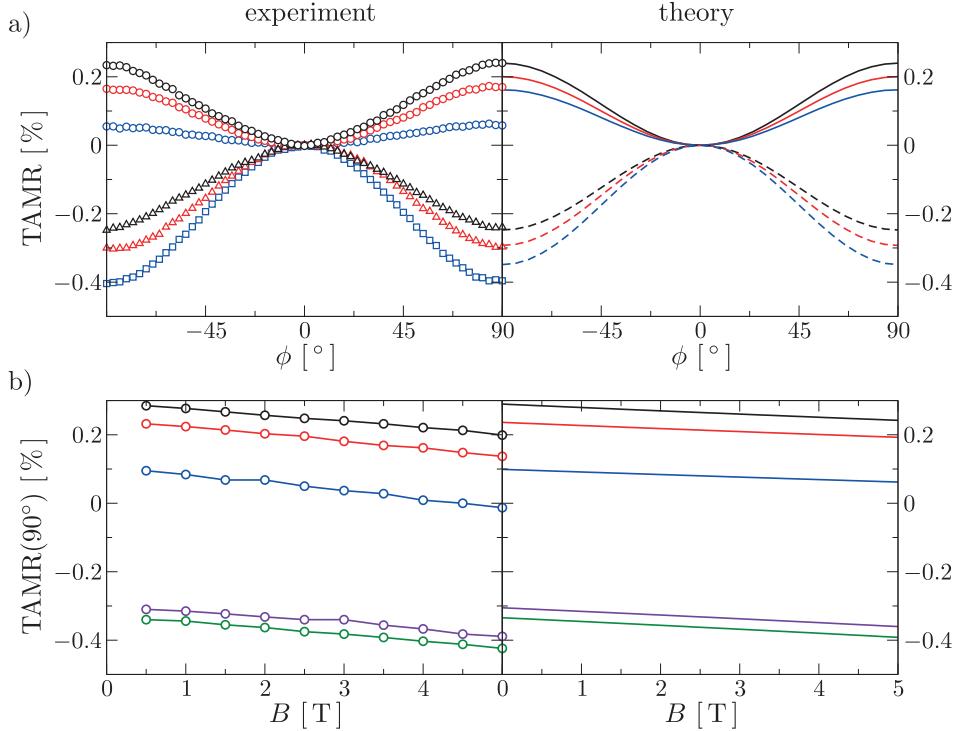


Figure 5.3 – (a) Angular dependence of the TAMR effect as found in the experiment (left panel, from [150]) and from numerical simulations (right panel). Data is shown for $V_{\text{bias}} = +90$ meV (experiment: \circ ; theory: $\alpha = 25.1 \text{ eV } \text{\AA}^2$, solid lines) and $V_{\text{bias}} = -90$ meV (experiment: \triangle ; theory: $\alpha = -25 \text{ eV } \text{\AA}^2$, dashed lines), as well as different magnetic fields: $B = 0.5$ T (black), $B = 5$ T (red), and $B = 10$ T (blue). (b) TAMR(90°) as a function of magnetic field in the experiment (left panel, from [48]) and from numerical simulations (right panel). Data is shown for $V_{\text{bias}} = +135$ meV ($\alpha = 29.3 \text{ eV } \text{\AA}^2$, black), $V_{\text{bias}} = +90$ meV ($\alpha = 24.3 \text{ eV } \text{\AA}^2$, red), $V_{\text{bias}} = +50$ meV ($\alpha = 10.5 \text{ eV } \text{\AA}^2$, blue), $V_{\text{bias}} = -50$ meV ($\alpha = -33.4 \text{ eV } \text{\AA}^2$, green), and $V_{\text{bias}} = -90$ meV ($\alpha = -30.8 \text{ eV } \text{\AA}^2$, violet).

orbit coupling [13, 151],

$$T(E, \mathbf{k}_{||}) = T^{(0)}(E, k_{||}) + T^{(1)}(E, k_{||}) \mathbf{n} \cdot \mathbf{w}(\mathbf{k}_{||}) + T^{(2)}(E, k_{||}) (\mathbf{n} \cdot \mathbf{w}(\mathbf{k}_{||}))^2 + \dots, \quad (5.14)$$

where

$$\mathbf{w}(\mathbf{k}_{||}) = \begin{pmatrix} \tilde{\alpha}k_y + \tilde{\gamma}k_x \\ -\tilde{\alpha}k_x - \tilde{\gamma}k_y \\ 0 \end{pmatrix} \quad (5.15)$$

is an effective spin-orbit field, obtained by averaging the spin-orbit field $\mathbf{B}_{\text{eff}}(z)$, $H_{\text{SO}} = H_{\text{D}} + H_{\text{BR}} = \mathbf{B}_{\text{eff}}(z) \cdot \boldsymbol{\sigma}$ over the unperturbed states of the system, i.e. in the absence of spin-orbit coupling. The effective Bychkov-Rashba and Dresselhaus parameters $\tilde{\alpha} = \alpha f_{\alpha}(k_{||})$ and $\tilde{\gamma} = \gamma f_{\gamma}(k_{||})$ are functions of $k_{||} = |\mathbf{k}_{||}|$ only, as are the expansion coefficients $T^{(n)}(E, k_{||})$. Note that the $T^{(n)}(E, k_{||})$ cannot be interpreted as transmission probabilities individually—for example, being expansion coefficients, $T^{(n)}(E, k_{||})$ may very well be negative.

As before, the total transmission is obtained by integrating over all transverse wave vectors,

$$T = T^{(0)} + T^{(1)} + T^{(2)} + \dots, \quad (5.16)$$

where $T^{(n)} = \langle T^{(n)}(E, k_{||}) (\mathbf{n} \cdot \mathbf{w}(\mathbf{k}_{||}))^n \rangle_{\mathbf{k}_{||}}$ and $\langle \dots \rangle_{\mathbf{k}_{||}} = \frac{1}{(2\pi)^2} \int d\mathbf{k}_{||} \dots$. In principle, the integration over transverse wave vectors must be restricted to values within the Fermi sphere. However, in a tunnel junction the transmission probability is suppressed exponentially for increasing $k_{||}$ (transmission is highest for perpendicular incidence, i.e. $\mathbf{k}_{||} = 0$), and hence we can extend the $\mathbf{k}_{||}$ -integration to infinity. To second order in the spin-orbit coupling, the conductance of the magnetic tunnel junction is then found as

$$G(\phi) = G^{(0)} + a_0^{(2)} + a_1^{(2)} \alpha \gamma \cos(2\phi), \quad (5.17)$$

where $G^{(0)}$, $a_0^{(2)}$, and $a_1^{(2)}$ are coefficients obtained by taking the respective $\mathbf{k}_{||}$ -integrations for the zeroth- and second-order term in Eq. (5.14) (for details, see Ref. [13]). In particular, the coefficient $a_1^{(2)}$ does not depend on the spin-orbit parameters α and γ . The linear term in Eq. (5.14) vanishes exactly upon integration, as $\mathbf{w}(-\mathbf{k}_{||}) = -\mathbf{w}(\mathbf{k}_{||})$.

After reviewing the phenomenological model for $B = 0$, we now turn to the case of finite magnetic field B . In this situation, we can still expand the transmission probability with respect to the spin-orbit coupling, albeit with different expansion coefficients and a different effective spin-orbit field:

$$T_B(E, \mathbf{k}_{||}) = T_B^{(0)}(E, \mathbf{k}_{||}) + T_B^{(1)}(E, \mathbf{k}_{||}) \mathbf{n} \cdot \mathbf{w}_B(\mathbf{k}_{||}) + T_B^{(2)}(E, \mathbf{k}_{||}) (\mathbf{n} \cdot \mathbf{w}_B(\mathbf{k}_{||}))^2 + \dots \quad (5.18)$$

In the following we will derive approximate relations valid to linear order in B between the expansion coefficients $T_B^{(n)}$ and the effective spin-orbit field \mathbf{w}_B to their counterparts for $B = 0$.

First, we consider the orbital effects of the magnetic field in the kinetic energy term of the Hamiltonian, and neglect the spin-orbit coupling. The kinetic part of the

Hamiltonian including a magnetic field reads

$$H_{\text{kin}} = \frac{\hbar^2}{2m} \left(\left(k_x - \frac{eB}{\hbar} \sin \theta z \right)^2 + \left(k_y + \frac{eB}{\hbar} \cos \theta z \right)^2 - \frac{\partial^2}{\partial z^2} \right), \quad (5.19)$$

where for the sake of argument we neglect the spatial dependence of the effective mass m . The effect of the magnetic field is two-fold: First, the maximum of the transmission probability is shifted with respect to $\mathbf{k}_{||} = 0$. Instead, the maximum of $T_B(E, \mathbf{k}_{||})$ is expected at $\mathbf{k}_{||,0}$ with

$$\left\langle \left(k_{x,0} - \frac{eB}{\hbar} \sin \theta z \right)^2 \right\rangle = 0 \quad \text{and} \quad \left\langle \left(k_{y,0} + \frac{eB}{\hbar} \cos \theta z \right)^2 \right\rangle = 0, \quad (5.20)$$

where $\langle \dots \rangle$ denotes an average over a quantum mechanical state. Hence, $\mathbf{k}_{||,0}$ can be written as

$$k_{x,0} = +b_1 B \sin \theta \quad (5.21a)$$

$$k_{y,0} = -b_1 B \cos \theta, \quad (5.21b)$$

where b_1 is a coefficient that depends on both $\langle z \rangle$ and $\langle z^2 \rangle$. This shift can be regarded as the effect of the Lorentz force on the electrons. Second, the overall transmission probability decreases [156]. However, this effect is quadratic in B and will thus be neglected in the following. These considerations motivate the approximation

$$T_B^{(n)}(E, \mathbf{k}_{||}) \approx T^{(n)} \left(E, \sqrt{(k_x - b_1 B \sin \theta)^2 + (k_y + b_1 B \cos \theta)^2} \right). \quad (5.22)$$

In addition to the transmission probability, also the effective spin-orbit field is shifted by the orbital terms from the magnetic field in Eqs. (5.6) and (5.7). The spin-orbit terms including a magnetic field read

$$H_{\text{SO}} = \alpha \left(\sigma_x \left(k_y + \frac{eB}{\hbar} \cos \theta z \right) - \sigma_y \left(k_x - \frac{eB}{\hbar} \sin \theta z \right) \right) \delta(z - z_l) + \gamma \left(\sigma_x \left(k_x - \frac{eB}{\hbar} \sin \theta z \right) - \sigma_y \left(k_y + \frac{eB}{\hbar} \cos \theta z \right) \right) \frac{\partial}{\partial z} \left(\gamma(z) \frac{\partial}{\partial z} \right), \quad (5.23)$$

and therefore the effective spin-orbit field is given as

$$\mathbf{w}_B(\mathbf{k}_{||}) \approx \mathbf{w}(k_x - b_2 B \sin \theta, k_y + b_2 B \cos \theta), \quad (5.24)$$

where b_2 is a coefficient that only depends on $\langle z \rangle$, as the spin-orbit terms only contain linear momenta in the transverse directions. Thus, in general we can expect $b_1 \neq b_2$.

With the approximations (5.22) and (5.24) we can now evaluate the total transmission. Since the zeroth order term is to first order only shifted by the magnetic field, we obtain the same result as in the case without magnetic field after integration:

$$\langle T_B^0(E, \mathbf{k}_{||}) \rangle_{\mathbf{k}_{||}} \approx T^{(0)} \quad (5.25)$$

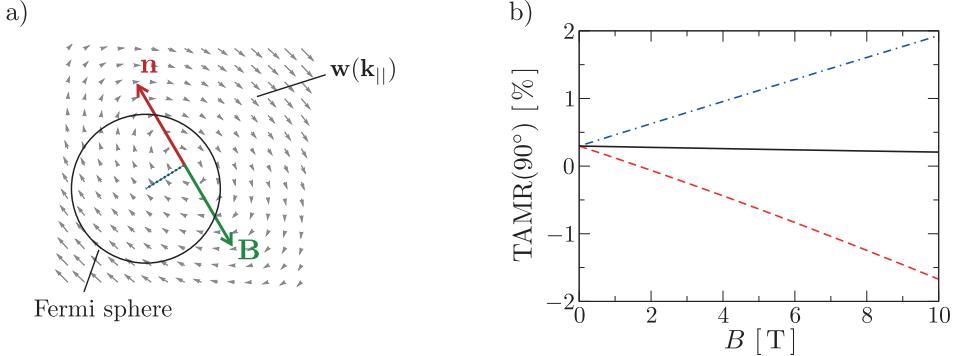


Figure 5.4 – (a) Schematic picture of the influence of the magnetic field: The Fermi sphere (or, more precisely, the maximum of the transmission probability) is shifted with respect to the effective spin-orbit field $\mathbf{w}(\mathbf{k}_{||})$. (b) Magnetic field dependence of TAMR(90°) when the magnetic vector potential is only included in the kinetic term (red dashed line), only in the spin-orbit term (blue dash-dotted line), or in both terms (black solid line).

The term linear in the spin-orbit coupling evaluates to

$$\begin{aligned}
 & \langle T_B^1(E, \mathbf{k}_{||}) \mathbf{n} \cdot \mathbf{w}_B(\mathbf{k}_{||}) \rangle_{\mathbf{k}_{||}} \\
 & \approx \frac{1}{(2\pi)^2} \int d\mathbf{k}_{||} T^{(1)}(E, k_{||}) \mathbf{n} \cdot \mathbf{w}(k_x + (b_1 - b_2)B \sin \theta, k_y - (b_1 - b_2)B \cos \theta) \\
 & = \frac{1}{(2\pi)^2} \int d\mathbf{k}_{||} T^{(1)}(E, k_{||}) B(b_1 - b_2) \\
 & \quad (-\tilde{\alpha} \cos^2 \theta + \tilde{\gamma} \sin \theta \cos \theta - \tilde{\alpha} \sin^2 \theta + \tilde{\gamma} \cos \theta \sin \theta) \\
 & = (b_1 - b_2)B (\langle T^{(1)}(E, k_{||}) f_\gamma(k_{||}) \rangle \gamma \sin(2\theta) - \langle T^{(1)}(E, k_{||}) f_\alpha(k_{||}) \rangle \alpha). \quad (5.26)
 \end{aligned}$$

In the second line we used the coordinate substitution $k_x \rightarrow k_x + b_1 B \sin \theta$ and $k_y \rightarrow k_y - b_1 B \cos \theta$, and in the third line we made use of the fact that terms linear in k_x or k_y vanish upon integration.

Hence, we find that the linear term, that vanished in the case without a magnetic field, gives a finite contribution for $B \neq 0$. This is due to the fact that the Fermi sphere is shifted by the Lorentz force. In particular, the effective spin-orbit field is shifted with respect to the center of the transmission maximum, as sketched in Fig. 5.4(a). In addition, we find a different *angular* dependence for the Bychkov-Rashba and the Dresselhaus spin-orbit coupling. This is due to different symmetries of the spin-orbit fields: The Bychkov-Rashba field exhibits a rotational symmetry [13] and thus leads to an angular-independent term in Eq. (5.26), whereas the Dresselhaus field does not, and hence leads to a angular dependence.

Finally, in addition to the original contribution for $B = 0$, the second-order term

in the spin-orbit coupling also contains terms quadratic in B , due to the shift of the Fermi sphere. As before, we neglect these terms and approximate

$$\langle T_B^2(E, \mathbf{k}_{||}) \rangle_{\mathbf{k}_{||}} \approx T^{(2)}. \quad (5.27)$$

Hence, the conductance of the magnetic tunnel junction in a magnetic field B takes the form

$$G_B(\phi) = G^{(0)} + a_0^{(2)} + a_2^{(1)} \alpha B + (a_1^{(2)} \alpha \gamma - a_1^{(1)} \gamma B) \cos(2\phi), \quad (5.28)$$

where the coefficients

$$a_1^{(1)} = -\frac{Se^2}{(2\pi)^2 h} \langle T^{(1)}(E, k_{||}) f_\gamma(k_{||}) \rangle (b_1 - b_2) \quad (5.29a)$$

$$a_2^{(1)} = -\frac{Se^2}{(2\pi)^2 h} \langle T^{(1)}(E, k_{||}) f_\alpha(k_{||}) \rangle (b_1 - b_2) \quad (5.29b)$$

do not depend on the spin-orbit parameters. The TAMR then evaluates to

$$\text{TAMR}(\phi) \approx \frac{G_B(0) - G_B(\phi)}{G^{(0)}} \propto (a_1^{(2)} \alpha \gamma - a_1^{(1)} \gamma B)(1 - \cos(2\phi)). \quad (5.30)$$

This equation gives a qualitative explanation of all characteristic features observed in the experiment: First, it reproduces the linear change in the TAMR with magnetic field. Second, the slope of this change is independent of the bias voltage, as it only depends on the (bulk) Dresselhaus parameter. Thus, all the experimentally observed features can be explained with the orbital effects caused by the magnetic field.

From the numerical results we can deduce that $a_1^{(2)}, a_1^{(1)} > 0$. In addition, the numerical results also confirm that $a_1^{(1)}$ contains a factor $(b_1 - b_2)$, i.e. opposing contributions from the orbital effects in the kinetic energy and in the spin-orbit coupling. In Fig. 5.4(b) we show the magnetic field dependence of $\text{TAMR}(90^\circ)$ when the magnetic vector potential is included only in the kinetic term (dashed line), only in the spin-orbit terms (dash-dotted line) and in both terms (solid line). When the magnetic field is only included in one term, we find a large change of the TAMR effect with magnetic field, however with opposite signs. When the vector potential is included in both the kinetic and the spin-orbit term, the contributions cancel to a large degree, leaving only a small effect. These numerical findings support the phenomenological model developed in this section.

Finally, it should be mentioned that although the magnetic field dependence of the TAMR (5.30) only depends on the Dresselhaus parameter γ , the conductance (5.28) also contains a B -dependent contribution from the Bychkov-Rashba parameter α . This term leads to a bias-voltage dependent linear slope in the magnetic field dependence of the conductance, in addition to the usual quadratic dependence [156]. Hence, it would be of great interest if such a dependence was indeed observed in the experiments.

5.3 Summary

In this chapter we have studied the magnetic field dependence of the TAMR effect in epitaxially grown Fe/GaAs/Au tunnel junctions. For this, we have extended the model used in previous works [47, 151, 152], which explained the TAMR effect through the spin-orbit coupling in the GaAs barrier, to include the orbital effects of a magnetic field. No additional fitting parameters were introduced in this step.

The observed magnetic field dependence—linear change in the TAMR with B and a bias-independent slope—could be explained both from the numerical simulations and a phenomenological model, where the orbital effects of the magnetic field were included as a shift of the Fermi sphere. In particular, the phenomenological model predicts that the slope of the TAMR change for a magnetic field B depends only on the (bulk) Dresselhaus spin-orbit parameter, and is hence bias-independent. This and other predictions from the phenomenological model are reproduced in numerical simulations, supporting its validity.

The TAMR effect in Fe/GaAs/Au tunnel junctions has been attributed to the effects of spin-orbit coupling in the GaAs barrier [47, 151, 152]. Including orbital effects into this model, we could explain the experimentally observed magnetic field dependence *without any additional fitting parameters*. This strongly supports the initial assumption that indeed spin-orbit coupling is the reason for the TAMR effect.

Chapter 6

The graphene edge state

6.1 Introduction

Graphene is a two-dimensional crystal of carbon atoms. Since its experimental discovery in 2004 [29], graphene has attracted a tremendous amount of interest (for reviews, see [31, 35, 157]). On the one hand, this is due to the fact that charge carriers in graphene obey an effective equation that mimics relativistic dynamics, leading to many unique physical properties, such as the odd-integer quantum Hall effect [32] that even prevails at room temperature [33], or Klein tunneling [34]. On the other hand, graphene is also a very promising material for electronics, because of its high mobility [36, 37], even at room-temperature [29], paving the road to ballistic transistors [31].

However, since (bulk) graphene does not feature a band gap, conventional transistor action is difficult to achieve. In contrast, graphene nanoribbons, strips of finite width cut from a graphene sheet, exhibit a finite band gap scaling with the nanoribbon width [158–160], and, as a consequence, room-temperature transistor action has been demonstrated [161]. The origin of this band gap is still under discussion [162–166].

Whereas the edges of graphene nanoribbons were not well-controlled in the early experiments, recent progress in fabrication has shown that controlled manufacturing of nanoribbons with atomically defined edges seems within experimental reach in the near future [160, 167, 168]. From a theoretical point of view, the electronic properties of nanoribbons are expected to depend crucially on the structure of the edge [169]. In particular, a zigzag or zigzag-like edge is expected to support a state localized at the graphene edge [169–171]. This edge state has been observed experimentally using scanning tunneling (STM) microscopy for a graphene edge on the surface of graphite¹ [172–174]. Due to electron-electron interactions, the edge state is expected to exhibit

¹ Up to now, a direct experimental confirmation of the edge state for exfoliated or epitaxially grown graphene is still missing.

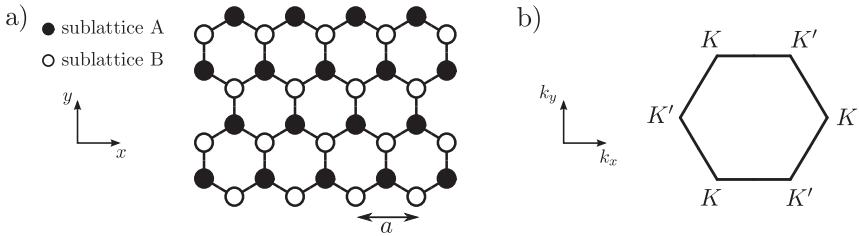


Figure 6.1 – (a) Lattice structure of graphene. (b) First Brillouin zone of graphene.

magnetism [170, 175, 176]. This magnetism can be used in edge state based spintronics that we present in Chapter 7.

On the theoretical side, the transport in situations, where the edge state is important, is almost exclusively studied in nearest-neighbor tight-binding approximation (for example, see Refs. [162, 165, 166, 177–181]). However, it has already been shown that the effects of next-nearest neighbor hopping on the edge state are significant in terms of the band structure [182, 183]. In addition, edge magnetism also changes the edge state related band structure considerably [164, 176].

Therefore, it is the intent of this chapter to study the fundamental properties of edge state transport in graphene. Doing so, we show that the paradigm model of transport in graphene, the nearest-neighbor tight-binding model, is not suitable for studying edge state transport: The results of edge transport in nearest-neighbor tight-binding approximation change fundamentally upon exponentially small corrections to this model.

To this end, the chapter is organized as follows: We begin by summarizing the main results of the tight-binding model of graphene employed in this work (for a complete discussion, see Appendix G). In addition, we review symmetries of the graphene Hamiltonian, that impose certain restrictions on the graphene wave function. These results are then applied to characterize the properties of the edge state for different approximations to the graphene Hamiltonian. Finally, we discuss extensively the transport properties of the graphene edge state when corrections to the nearest-neighbor tight-binding approximation are included.

6.2 Fundamentals of graphene

6.2.1 Graphene Hamiltonian

The electronic structure of graphene is usually described in terms of a tight-binding model, that is discussed extensively in Appendix G. Here, we merely summarize the main results.

Graphene is a two-dimensional sheet of carbon atoms, arranged in a honeycomb network, as shown in Fig. 6.1(a). The honeycomb network is not a primitive Bravais lattice, but is generated by a triangular lattice with two basis atoms. These basis atoms are labeled A and B, respectively, and the triangular lattice constant is denoted as a .

In single-orbital tight-binding approximation, the graphene Hamiltonian reads

$$H = \sum_{i,\alpha,j,\beta} t_{i\alpha j\beta} |i, \alpha\rangle \langle j, \beta| . \quad (6.1)$$

Here, $|i, \alpha\rangle$ denotes the orbital at lattice site $\mathbf{R}_i + \mathbf{d}_\alpha$, where \mathbf{R}_i is a (triangular) lattice vector and \mathbf{d}_α the relative position of atom $\alpha = \text{A}, \text{B}$ within the unit cell. We consider matrix elements up to next-nearest neighbors, so that

$$t_{i\alpha j\beta} = \begin{cases} V_i + M_i & \text{if } i = j, \alpha = \beta = \text{A} \\ V_i - M_i & \text{if } i = j, \alpha = \beta = \text{B} \\ -t & \text{if } i, \alpha \text{ and } j, \beta \text{ are nearest neighbors,} \\ -t' & \text{if } i, \alpha \text{ and } j, \beta \text{ are next-nearest neighbors.} \end{cases} \quad (6.2)$$

Here V_i is a potential that has the same value on both sublattices, and M_i is a staggered potential, i.e. has opposite sign on atom A and B, respectively. The numerical values of the hopping parameter are chosen as $t = 2.7 \text{ eV}$ and $t' = 0.1t$ [183, 184].

In the case of an infinite graphene sheet and constant potentials $V_i = V_0$ and $M_i = M$, Bloch's theorem holds and in \mathbf{k} -space the Hamiltonian reads (Eq. (G.23))

$$H(\mathbf{k}) = \left(V_0 - 2t' \sum_{l=1}^3 \cos(\mathbf{k}\delta'_l) \right) \mathbb{1} - \left(t \sum_{l=1}^3 \cos(\mathbf{k}\delta_l) \right) \sigma^x + \left(t \sum_{l=1}^3 \sin(\mathbf{k}\delta_l) \right) \sigma^y + M\sigma^z , \quad (6.3)$$

where $\sigma^{x,y,z}$ are the Pauli matrices and δ_l (δ'_l) the distance vectors between nearest (next-nearest) neighbors. Note that the two basis atoms in the unit cell give rise to a wave function with two components, referring to sublattice A and B,

$$\varphi(j) = \begin{pmatrix} \varphi_{\text{A}}(j) \\ \varphi_{\text{B}}(j) \end{pmatrix} \quad (6.4)$$

such that the Hamiltonian $H(\mathbf{k})$ can be written as a 2×2 -matrix in Eq. (6.3).

The wave vector \mathbf{k} is defined within the first Brillouin zone of graphene that has the shape of a hexagon, as shown in Fig. 6.1(b). Only two of the six corners of the Brillouin zone are unique, as always three corners are connected by a reciprocal lattice vector. These two non-equivalent corners are called K and K' -point, respectively. Here we choose

$$\mathbf{K}_\tau = \begin{pmatrix} \tau \frac{4\pi}{3a} \\ 0 \end{pmatrix} \quad (6.5)$$

to represent the K ($\tau = +1$, $\mathbf{K} = \mathbf{K}_+$) and K' -point ($\tau = -1$, $\mathbf{K}' = \mathbf{K}_-$).

In a system close to charge neutrality, the states at the Fermi energy E_F have a wave vector close to the K and K' -point. States close to \mathbf{K} and \mathbf{K}' are said to be in the K and K' -valley. In the low energy regime and for smoothly varying potentials, the particle dynamics can be described by an effective Hamiltonian in each valley (Eq. (G.36)),

$$H_\tau = -i\hbar v_F (\tau \partial_x \sigma^x + \partial_y \sigma^y) + V(\mathbf{x}) + M(\mathbf{x}) \sigma^z, \quad (6.6)$$

where $H_K = H_+$ is the effective Hamiltonian for the K -valley and $H_{K'} = H_-$ for the K' -valley. Here, $V(\mathbf{x})$ and $M(\mathbf{x})$ are a smoothly varying potential and staggered potential, respectively, and $v_F = \sqrt{3}ta/2 \approx 10^6$ m/s is the Fermi velocity. Again, the two sublattices give rise to a two-component envelope wave function

$$\varphi(\mathbf{x}) = \begin{pmatrix} \varphi_A(\mathbf{x}) \\ \varphi_B(\mathbf{x}) \end{pmatrix}, \quad (6.7)$$

and the Pauli matrices $\sigma^{x,y,z}$ act within this sublattice space.

The effective Hamiltonian in Eq. (6.6) is equivalent to a Dirac Hamiltonian in two spatial dimensions, albeit with an “effective speed of light” v_F . As a consequence, the low-energy spectrum of graphene is linear close to the K and K' -point. This gives rise to a cone-like shape of the band structure in the K and K' -valley, that is commonly referred to as *Dirac cone*. Furthermore, the two-component wave function (6.7) behaves like a spinor under transformations, and the sublattice degree of freedom is thus referred to as *pseudospin*.

The envelope wave function (6.7) and the tight-binding wave function (6.4) are connected through (Eq. (G.38))

$$\varphi_\alpha(j) = e^{i\mathbf{K}(\mathbf{R}_j+\mathbf{d}_\alpha)} \varphi_\alpha(\mathbf{R}_j + \mathbf{d}_\alpha) + e^{i\mathbf{K}'(\mathbf{R}_j+\mathbf{d}_\alpha)} \varphi'_\alpha(\mathbf{R}_j + \mathbf{d}_\alpha), \quad (6.8)$$

where $\varphi_\alpha(\mathbf{x})$ and $\varphi'_\alpha(\mathbf{x})$ are the envelope wave functions in the K and K' -valley, respectively.

6.2.2 Symmetries of the graphene Hamiltonian

We will now review some important symmetries of the graphene Hamiltonian, in particular *particle-hole symmetries*. If the Hamiltonian H is particle-hole symmetric, there is a transformation \mathcal{P} such that

$$\mathcal{P}^\dagger H \mathcal{P} = -H. \quad (6.9)$$

Then, if φ is an eigenstate of H with energy E , $H\varphi = E\varphi$, $\mathcal{P}\varphi$ is an eigenstate with energy $-E$, $H\mathcal{P}\varphi = -E\mathcal{P}\varphi$. Thus, the energy spectrum is symmetric with respect to $E = 0$ and every “particle” wave function φ with positive energy has a “hole” partner $\mathcal{P}\varphi$ with negative energy.

Writing the tight-binding Hamiltonian in terms of Pauli matrices as in Eq. (6.3) has the advantage that we can treat it on equal footing with the effective Dirac Hamiltonian

(6.6). The symmetries and their consequences considered below are thus valid for both the tight-binding and the effective Hamiltonian.

The Hamiltonian is particle-hole symmetric under the unitary transformation

$$\mathcal{P}_z = \sigma^z, \quad (6.10)$$

if it contains only terms proportional to σ^x or σ^y . That is, \mathcal{P}_z is a particle-hole transformation if the Hamiltonian does not include potentials² $V(\mathbf{x})$ and $M(\mathbf{x})$ and, in the case of the tight-binding Hamiltonian, only includes nearest-neighbor hopping³.

In order to exist also in a system of finite extent, the particle-hole symmetry must additionally be compatible with the boundary conditions [171]. In a finite graphene patch, the wave function must vanish on a set of missing atoms, as discussed in Section G.3.2. Since \mathcal{P}_z does not mix sublattice A and B, it is compatible with any boundary.

The particle-hole transformation \mathcal{P}_z is especially important, as it also has consequences for the occupation probabilities of the two sublattices [185]. Note that $\frac{1}{2}(1 \pm \sigma^z)$ is a projection operator onto the wave function of sublattice A and B, respectively. From this projection operator we obtain the following expression for an eigenstate φ with energy E , provided that $\{\mathcal{P}_z, H\} = 0$:

$$\begin{aligned} \langle \varphi | \frac{1}{2}(1 + \sigma^z) H \frac{1}{2}(1 - \sigma^z) | \varphi \rangle &= \langle \varphi | \frac{1}{2}(1 + \sigma^z) \frac{1}{2}(1 + \sigma^z) H | \varphi \rangle \\ &= E \int d\mathbf{x} |\varphi_A(\mathbf{x})|^2 \end{aligned} \quad (6.11)$$

Commuting the Hamiltonian H to the left we also obtain

$$\langle \varphi | \frac{1}{2}(1 + \sigma^z) H \frac{1}{2}(1 - \sigma^z) | \varphi \rangle = E \int d\mathbf{x} |\varphi_B(\mathbf{x})|^2 \quad (6.12)$$

and thus, combining Eqs. (6.11) and (6.12),

$$\int d\mathbf{x} |\varphi_A(\mathbf{x})|^2 = \int d\mathbf{x} |\varphi_B(\mathbf{x})|^2 \quad \text{for } E \neq 0. \quad (6.13)$$

Hence, if the Hamiltonian is particle-hole symmetric under \mathcal{P}_z , every eigenstate with energy $E \neq 0$ occupies sublattice A and B with equal probability.

\mathcal{P}_z is not the only particle-hole symmetry in the system. For example, the effective Hamiltonian is particle-hole symmetric under the antunitary transformation [186]

$$\mathcal{P}_x = \mathcal{C}\sigma^x, \quad (6.14)$$

² Note that a constant potential $V(\mathbf{x}) = V_0$ only leads to an overall energy shift; as such, it does not break particle-hole symmetry.

³ In the low-energy regime, the next-nearest neighbor hopping does not play a role (see Section G.2.3). Thus, close to the Dirac point, the spectrum is still approximately particle-hole symmetric, even if next-nearest neighbor hopping is included. However, this is not true anymore, if edge states are present (see Section 6.3.4).

	σ^x, σ^y	$M(\mathbf{x})\sigma^z$	$V(\mathbf{x})\mathbf{1}$	$t' \neq 0$
particle-hole symmetry	yes	yes ^a	no	no
$\int d\mathbf{x} \varphi_A(\mathbf{x}) ^2 = \int d\mathbf{x} \varphi_B(\mathbf{x}) ^2$	yes ^b	no	no	no

Table 6.1 – Summary of which terms in the graphene Hamiltonian break particle-hole symmetry, and which terms imply equal occupation probability on sublattice A and B.

^a In bulk; for armchair boundaries; for zigzag boundaries provided that there is an additional symmetry connecting zigzag edges of the opposite kind (such as in nanoribbons).

^b For $E \neq 0$.

where \mathcal{C} denotes complex conjugation. The tight-binding Hamiltonian, in turn, is particle-hole symmetric under

$$\mathcal{P}_y = \mathcal{C}\sigma^y. \quad (6.15)$$

Both \mathcal{P}_x and \mathcal{P}_y are particle-hole transformations if the respective Hamiltonian only contains terms proportional to σ^x , σ^y and σ^z , i.e. including a staggered potential.

However, unlike \mathcal{P}_z , these particle-hole symmetries interchange sublattice A and B and are thus only compatible with boundary conditions that involve both sublattices on equal footing, such as armchair boundary conditions. In particular, a zigzag edge breaks these particle hole symmetries—unless there is an additional symmetry connecting different zigzag edges as in a zigzag nanoribbon. Still, even if $\mathcal{P}_{x,y}$ is a particle-hole transformation, it does *not* give rise to a relation for the occupation probability of the two sublattices.

The findings of this section are summarized in Table 6.1.

6.3 Characterization of the edge state

6.3.1 Edge state at a single zigzag edge

Over the course of the next sections, we will examine the graphene edge state under different conditions. Here, we start with the simplest situation, a half-infinite graphene sheet with a single zigzag edge in nearest-neighbor tight-binding approximation.

It has been known for some time that a zigzag edge in a honeycomb network supports a state localized at the edge [96, 187], but it was Fujita that first noticed the physical relevance of this state [170]. As Fujita, we consider a half-infinite graphene sheet with a single zigzag edge in x -direction, as sketched in Fig. 6.2(a). In this situation, the outermost atoms are all of sublattice B, and the wave function vanishes on the first row of missing A atoms.

The half-infinite graphene sheet can be understood as a repetition of unit cells of length a , where the unit cell is given as a chain of carbon atoms, as indicated in black in Fig. 6.2(a). We label atoms within this unit cell by (j, α) , where $j = 0, 1, 2, \dots$ increases when moving away from the edge and $\alpha = A, B$ denotes the sublattice. Note that in

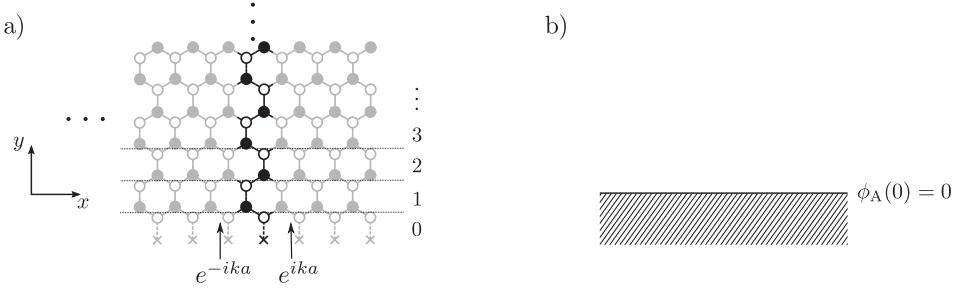


Figure 6.2 – (a) A half-infinite graphene sheet with a single zigzag edge in nearest-neighbor tight-binding approximation. The unit cell of the system is indicated in black, and the first row of missing atoms is indicated by crosses. (b) The same situation within the effective Dirac Hamiltonian.

our convention, the first missing atom has index $(0, A)$. The unit cells themselves are labeled with an index n , that increases with increasing x .

Because of translational symmetry in x -direction, we can write the wave function of the system in Bloch form,

$$\varphi_\alpha(n, j) = \tilde{\phi}_\alpha(j) e^{i k a n}, \quad (6.16)$$

where $k \in [-\frac{\pi}{a}, \frac{\pi}{a}]$. Within a unit cell, the atomic positions change by $\pm \frac{a}{2}$ in x -direction when going from j to $j + 1$ or $j - 1$. In order to eliminate the phase associated with this lateral displacement, it is convenient to define

$$\tilde{\phi}_\alpha(j) = \phi_\alpha(j) \quad \text{for } j \text{ even,} \quad (6.17)$$

$$\tilde{\phi}_\alpha(j) = \phi_\alpha(j) e^{-i k a / 2} \quad \text{for } j \text{ odd.} \quad (6.18)$$

Making use of the Bloch form (6.16) we can then write down the stationary Schrödinger equation $H\varphi = E\varphi$ as a set of equations only involving $\phi_\alpha(j)$:

$$E \phi_A(j) = -t \phi_B(j) + t e^{-\gamma} \phi_B(j - 1) \quad (6.19)$$

$$E \phi_B(j) = -t \phi_A(j) + t e^{-\gamma} \phi_A(j + 1), \quad (6.20)$$

where we defined $e^{-\gamma} = -2 \cos(ka/2)$. This set of equations is closed by the boundary condition $\phi_A(0) = 0$, so that

$$E \phi_B(0) = -t \phi_A(1). \quad (6.21)$$

It is easy to see that these equations have a solution for $E = 0$:

$$\phi_A(j) = 0, \quad \phi_B(j) = e^{-\gamma j}. \quad (6.22)$$

This solution is normalizable—and thus represents a physical solution—if $|e^{-\gamma}| = |2 \cos(ka/2)| < 1$, i.e. $-\frac{\pi}{a} \leq k < -\frac{2\pi}{3a}$ and $\frac{2\pi}{3a} < k < \frac{\pi}{a}$. From

$$\sum_{j=0}^{\infty} e^{-2\gamma j} = \frac{1}{1 - e^{-2\gamma}} \quad (6.23)$$

we find the normalized solution

$$\phi_A(j) = 0 \quad \phi_B(j) = \sqrt{1 - e^{-2\gamma}} e^{-\gamma j}. \quad (6.24)$$

This is the celebrated graphene edge state. Note that this state resides on a single sublattice—since it is a zero-energy solution, Eq. (6.13) does not hold, although the nearest-neighbor tight-binding Hamiltonian is particle-hole symmetric with respect to \mathcal{P}_z .

The edge state is a solution that decays exponentially in the direction perpendicular to the edge. The perpendicular decay constant γ , and thus the perpendicular extension of the edge state, depends uniquely on the lateral wave vector k , $\gamma = \gamma(k)$. This is in contrast to an edge in an ordinary, semiconductor based two-dimensional electron gas, where the perpendicular and the lateral direction are decoupled. The edge state is maximally localized for $k = \pm \frac{\pi}{a}$, where it resides on the outermost atom only,

$$|\phi_B(j)|^2 = \begin{cases} 1 & \text{for } j = 0 \\ 0 & \text{else,} \end{cases} \quad (6.25)$$

and extends through the whole system for $k = \mp \frac{2\pi}{3a}$ (that is, at the K and K' -point, respectively),

$$|\phi_B(j)|^2 = \text{const.} \quad (6.26)$$

Close to the K -points, we find

$$\gamma = i\pi - \frac{\sqrt{3}a}{2} \left(k \mp \frac{2\pi}{3a} \right) + \mathcal{O} \left(\left(k \mp \frac{2\pi}{3a} \right)^2 \right), \quad (6.27)$$

i.e. the perpendicular decay constant depends linearly on the lateral wave vector k .

In contrast to the decay length, the energy of the edge state $E(k) = 0$ is a constant, independent of k . Thus, the edge state is *dispersionless*, i.e. has zero group velocity. In Fig. 6.3 we show a schematic picture of the band structure of a half-infinite graphene sheet including the edge state (for a discussion on how to project the two-dimensional band structure of graphene into one dimension, see Section G.3.1). The edge state band is perfectly flat and connects the two K -points.

After discussing the edge state within the tight-binding model, we now turn to the effective Dirac Hamiltonian. Again, because of translational symmetry in x -direction, the wave function of the system can be written in Bloch form

$$\varphi(\mathbf{x}) = \phi(y) e^{iqx}. \quad (6.28)$$

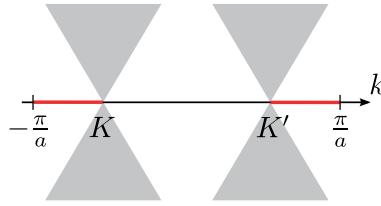


Figure 6.3 – Schematic depiction of the edge state dispersion $E(k)$, shown in red. The Dirac cones are indicated in light grey.

Note that a wave vector in the effective Hamiltonian is always given with respect to the K or K' -point, as discussed in Section G.2.3. We will always denote such a relative wave vector by q , in order to distinguish it from a wave vector k in the tight-binding model.

Using Eq. (6.28) in the effective Dirac Hamiltonian (6.6) we obtain a one-dimensional differential equation for the transverse wave function

$$\hbar v_F (\tau q - \partial_y) \phi_B(y) = E \phi_A(y) \quad (6.29)$$

$$\hbar v_F (\tau q + \partial_y) \phi_A(y) = E \phi_B(y), \quad (6.30)$$

where $\tau = +1$ for the K -valley and $\tau = -1$ for the K' -valley. The boundary condition for a half-infinite graphene sheet as sketched in Fig. 6.2(b) is then

$$\phi_A(0) = 0. \quad (6.31)$$

For $E = 0$, these equations have the solution

$$\phi_A(y) = 0 \quad \phi_B = e^{\tau q y}. \quad (6.32)$$

This solution is normalizable, if $\tau q < 0$, i.e. for negative (positive) wave vectors q with respect to the $K(K')$ -point. Note that this is consistent with the tight-binding result, where $k < -\frac{2\pi}{3a}$ (K -point) or $k > \frac{2\pi}{3a}$ (K' -point). The normalized edge state solution then reads

$$\phi(y) = \begin{pmatrix} 0 \\ \sqrt{-2\tau q} e^{\tau q y} \end{pmatrix}. \quad (6.33)$$

Thus, also within the effective Hamiltonian we find an edge state solution that decays exponentially in the perpendicular direction. Again, the decay length depends on the lateral wave vector q . Furthermore, the perpendicular decay constant τq is in agreement with the linear expansion of γ close to the K -points, Eq. (6.27).⁴ Thus, close to the K -points, the edge state is well-described by the effective Dirac Hamiltonian.

⁴ For this, note that the y -coordinate of the j -th atom is given by $y_j = j\sqrt{3}a/2$.

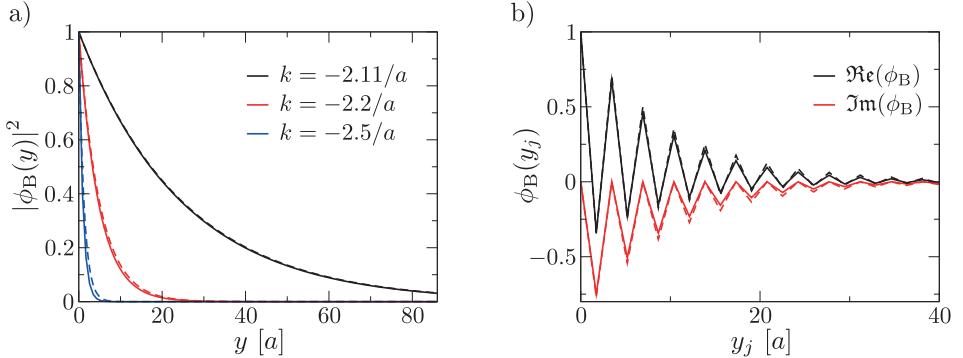


Figure 6.4 – Comparison between the edge state solutions of the tight-binding (solid lines) and the effective Hamiltonian (dashed lines). (a) shows the probability density $|\phi_B(y)|^2$ for different values of k , and (b) the wave functions $\phi_B(y)$ for $k = -2.3/a$ as a function of the distance y from the edge. In (b), we denote this distance as y_j in order to emphasize that the values of the wave function are only given for fixed atom positions. In order to conveniently compare the wave functions for different values of k , they have been normalized such that $\phi_B(0) = 1$.

In Fig. 6.4(a) we show a comparison between the edge state from the tight-binding and the effective Hamiltonian for different values of k . As expected, close to the K -point, both solutions agree very well. Moving away from the K -point, the edge state from the effective Hamiltonian decays somewhat slower than the edge state within the tight-binding model. In fact, the tight-binding edge state solution is maximally localized for $k = \pm\frac{\pi}{a}$, whereas the edge state solution in the effective Hamiltonian is localized strongest for $\tau q \rightarrow -\infty$. The point of maximal localization, $k = \pm\frac{\pi}{a}$, is located halfway between the K and K' -points. Around this point, the validity of the effective Hamiltonian, that is an expansion around a *single* K -point, breaks down. Thus, it is not surprising that the effective Hamiltonian underestimates the strength of the edge state decay, but as seen in Fig. 6.4, this deviation is small.

In addition to the exponential decay, the edge state also features a change in sign from atom to atom, since $e^{-\gamma} < 0$. In order to obtain such an oscillation on the scale of the lattice from the smooth envelope wave functions of the effective Hamiltonian, we have to use Eq. (6.8) to relate the envelope wave functions to the tight-binding solution. For the case of the edge state, this is shown in Fig. 6.4(b), where we find a good agreement between both solutions.

Hence, the effective Dirac Hamiltonian describes the main features of the graphene edge state rather well, but somewhat underestimates the strength of the decay. Therefore, when keeping in mind this limitation, the effective Hamiltonian can be applied successfully to study the edge state in more complex situations, when an analytical solution of the tight-binding model is not possible anymore.

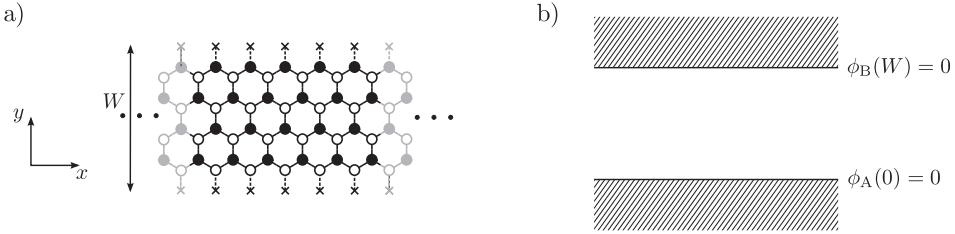


Figure 6.5 – (a) A zigzag graphene nanoribbon and (b) its equivalent within the effective Dirac Hamiltonian.

6.3.2 Edge state in a zigzag nanoribbon

We now turn to a system containing *two* parallel zigzag edges, i.e. a zigzag graphene nanoribbon. It is still possible to solve the tight-binding Hamiltonian in nearest-neighbor approximation analytically in this case [183]. However, the resulting expressions are rather unwieldy. Thus, here we restrict the discussion to the effective Dirac Hamiltonian in order to obtain the characteristic features of the edge state in a nanoribbon.

Since the zigzag nanoribbon again has translational symmetry in x -direction, the eigenstates will have the Bloch form as given in Eq. (6.28). The boundary conditions then read

$$\phi_A(0) = 0, \quad \phi_B(W) = 0, \quad (6.34)$$

where W is the width of the nanoribbon (see Fig. 6.5). The problem of solving the effective Dirac Hamiltonian (6.6) with these boundary conditions has been addressed in Refs. [53, 185]. The edge state solution is given as

$$\phi(y) \cong \begin{pmatrix} \sinh(zy) \\ \operatorname{sgn}(E) \sinh(z(W-y)) \end{pmatrix} \quad (6.35)$$

where the decay constant z is real and obeys

$$q = -\tau \frac{z}{\tanh(zW)}. \quad (6.36)$$

Again, $\tau = +1$ for the K and $\tau = -1$ for the K' -valley. The edge state has the eigenenergy

$$E = \pm \frac{z}{\sinh zW}. \quad (6.37)$$

The solution (6.35) is decaying exponentially in the perpendicular direction and can thus be called an edge state, if the transcendental equation (6.36) has a real solution z . This is the case for $\tau q < -1/W$. Note that the critical value of q for obtaining an edge state solution has been shifted slightly with respect to the single edge, where

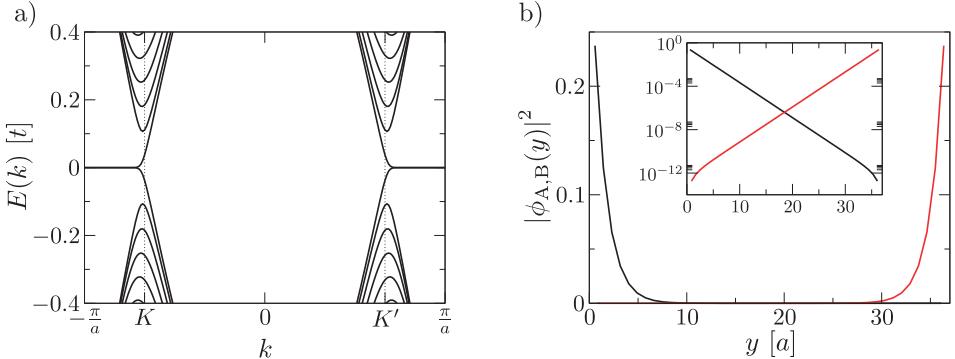


Figure 6.6 – A zigzag nanoribbon in nearest-neighbor tight-binding approximation: (a) The band structure of the zigzag nanoribbon. (b) Probability density profile of the edge state across the zigzag nanoribbon, for $k = -2.4/a$. The inset shows the same density profile on a logarithmic scale. The probability density for sublattice A is shown in red, for sublattice B in black. In both (a) and (b) the nanoribbon had a width $W = 64a/\sqrt{3}$.

$\tau q < 0$ was required for the existence of an edge state. For wide enough ribbons, this difference becomes negligible.

The state at $q = -\tau/W$ has $z = 0$ and connects edge state and bulk-like solutions, as does the $q = 0$ state for the single edge. However, unlike for the case of the single edge, where $\phi_B(y) = \text{const.}$, this connecting state decreases linearly from one edge to the other [53], as it must satisfy the boundary conditions at *both* edges.

For $zW \gg 1$, i.e. a very wide ribbon, or a strongly decaying state, from Eq. (6.36) we obtain $z \approx -\tau q$ and the eigenstate simplifies to

$$\phi(y) \approx \begin{pmatrix} e^{\tau q(W-y)} \\ \pm e^{\tau qy} \end{pmatrix}. \quad (6.38)$$

In this limit, the edge state in the zigzag nanoribbon is given as two decaying exponentials at each edge, residing on opposite sublattices (the boundary conditions at both edges are opposite to each other), reminiscent of the edge state at a single edge as discussed in the previous section. From Eq. (6.37) we find the energy of this state as

$$E \approx \mp 2\tau q e^{\tau q W}. \quad (6.39)$$

This energy is exponentially close to zero, since $-\tau q W = zW \gg 1$. Thus, we find two states with an exponentially small, but finite energy splitting. There are two states, since the nanoribbon features two edges.

These states appear as two almost dispersionless bands in the band structure of a zigzag nanoribbon as shown in Fig. 6.6(a), where we calculated numerically the band structure of a zigzag nanoribbon in nearest-neighbor tight-binding approximation.

Since the energy of these states is exponentially small, these bands appear to have the same energy when plotted in a linear scale. However, there is always an exponentially small, but finite energy splitting between these states at a given value of k , with the exception of the two zero-energy states at $k = \pm\frac{\pi}{a}$. There the effective Dirac theory breaks down, as seen in the previous section. Moving closer to the K -points, the splitting becomes larger, until the edge states join the extended, bulk-like states in the Dirac cones (see also Section G.3.1). The existence of this splitting has important consequences that we discuss below. Note that the edge state band merges smoothly into a band of bulk states: The states for $\tau q > -1/W$ are no longer localized to the edge, but extend through the whole width of the ribbon.

In Fig. 6.6(b) we show the probability density profile of an edge state across the nanoribbon. For the chosen value of the wave vector k , the edge state is localized to the edges, and resides on a single sublattice close to the edges, as expected from Eq. (6.38). The inset of Fig. 6.6(b) shows the probability density of the edge state in a logarithmic scale, where the exponential decay is more apparent. Close to the edges, we observe a small “kink” in the decay, as the state must also satisfy the boundary condition on the opposite edge.

A picture like Fig. 6.6(b) might lead to the impression that the two edges in the zigzag nanoribbon carry two independent edge states. This is not true: Fig. 6.6(b) shows a *single eigenstate* of the nanoribbon. The nanoribbon in nearest-neighbor tight-binding approximation is particle-hole symmetric under transformation with \mathcal{P}_z , and thus every eigenstate with nonzero energy must occupy both sublattices with equal probability, as discussed in Section 6.2.2. *It is thus not possible to find an eigenstate with nonzero energy that resides at a single edge and a single sublattice only.* Thus, any edge state with $k \neq \pm\frac{\pi}{a}$ is located with equal probability at both edges. Since the probability densities at both edges arise from a single state, they are inevitably coupled: whatever happens at one edge, will also influence the probability density at the other edge. This must be kept in mind when discussing the edge state in a zigzag nanoribbon⁵.

Note that the edge state solutions (6.35) and (6.38) are of the form

$$\begin{pmatrix} \phi_A(y) \\ 0 \end{pmatrix} \pm \begin{pmatrix} 0 \\ \phi_B(y) \end{pmatrix}, \quad (6.40)$$

i.e. a “bonding” and “anti-bonding” linear combination of wave functions on a single sublattice. Combining “bonding” and “anti-bonding” solutions, it is in fact possible to construct a wave function localized at single edge only. However, since these states have positive and negative energy respectively, the resulting wave function is not an eigenstate of the system.

Within the effective Dirac Hamiltonian, the “bonding” (“anti-bonding”) solution has a positive (negative) energy. In order to obtain the lattice wave function from the envelope wave functions, we again have to employ (6.8). The lattice structure

⁵ In fact, calling this state “edge state” is already a possible source of confusion.

of the zigzag nanoribbon then leads to a peculiar even-odd effect with respect to the nanoribbon width [188], that is of crucial importance in valleytronics [188, 189], but does not play a role in this work.

As seen in this section, due to the symmetries of the graphene Hamiltonian it is not possible to find a state localized at a single edge in a zigzag nanoribbon in nearest-neighbor tight-binding approximation⁶. In the next sections we therefore examine the behavior of the edge state including additional terms in the Hamiltonian that break this symmetry.

6.3.3 Effects of a staggered potential

We first consider the effect of a staggered potential M . In a half-infinite graphene sheet, the edge state resides on a single sublattice only. The staggered potential then simply shifts the edge state energy by $-M$, if the edge state resides on sublattice B, and $+M$, if the edge state resides on sublattice A.

In a nanoribbon, the situation is more complex, as the edge state is in principle located at both edges and both sublattices, as discussed in the previous section. However, since a staggered potential breaks the particle-hole symmetry \mathcal{P}_z , the eigenstates of the zigzag nanoribbon need not occupy both sublattices equally, and the system can localize to a single edge, with the energy shifted by $\pm M$, as we show within the framework of the effective Dirac Hamiltonian below.

The analysis of Ref. [53] that lead to the results presented in the previous section carries over straightforwardly to a zigzag nanoribbon with a constant staggered potential. We find the edge state solution as

$$\phi(y) \cong \begin{pmatrix} \sqrt{\frac{E+M}{E}} \sinh(zy) \\ \text{sgn}(E+M) \sqrt{\frac{E-M}{E}} \sinh(z(W-y)) \end{pmatrix}, \quad (6.41)$$

where the decay constant z is again a solution of

$$q = -\tau \frac{z}{\tanh(zW)}. \quad (6.42)$$

Thus, decaying states exist for $q < -\tau/W$, as before. The energy of these states is then given by

$$E = \pm \hbar v_F \sqrt{\left(\frac{M}{\hbar v_F}\right)^2 + \left(\frac{z}{\sinh zW}\right)^2}. \quad (6.43)$$

Hence, we find edge state solutions for energies $|E| \geq |M|$. Thus, the staggered potential opens a band gap of $2|M|$ for the edge state, just as for the bulk states (see Section

⁶ As discussed above, this is not true for the zero-energy state at $k = \pm \frac{\pi}{a}$. However, since this is a single state, it will only provide a negligible contribution when we are studying transport in later sections.

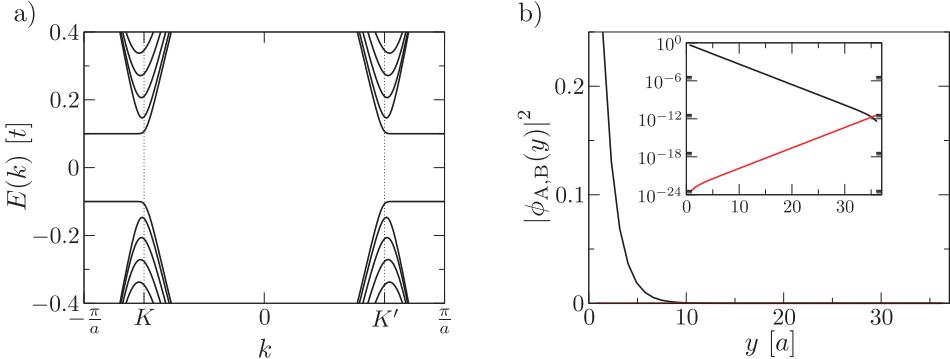


Figure 6.7 – A zigzag graphene nanoribbon in nearest-neighbor tight-binding approximation including a staggered potential $M = 0.1t$. (a) shows the band structure of the ribbon and (b) the probability density profile of the edge state across the nanoribbon, for $k = -2.4/a$ and $E \approx -M$. The inset shows the same density profile on a logarithmic scale. The probability density for sublattice A is shown in red, for sublattice B in black. In both (a) and (b) the nanoribbon had a width $W = 64a/\sqrt{3}$.

G.2.2). For $zW = -\tau qW \gg 1$ the energy dispersion simplifies to

$$E \approx \pm \hbar v_F \sqrt{\left(\frac{M}{\hbar v_F}\right)^2 + 4q^2 e^{-2\tau qW}}. \quad (6.44)$$

which therefore is exponentially close to $\pm|M|$.

In Fig. 6.7(a) we show the band structure of a zigzag graphene nanoribbon in nearest-neighbor tight-binding approximation with a constant staggered potential M . As predicted from the effective Hamiltonian, the staggered potential opens a band gap of size $2|M|$, so that we find two sets of almost dispersionless bands close to $E = \pm|M|$.

Note that the functional form of the edge state in a staggered potential (6.41) is the same as for the edge state with $M = 0$, Eq. (6.35), but the wave functions for sublattice A and B are multiplied with different weights: For $E \approx M$, sublattice B is suppressed, whereas for $E \approx -M$ sublattice A is suppressed. Since $|E|$ is exponentially close to $|M|$ for $-\tau qW \gg 1$, the respective sublattices are also suppressed exponentially in this case. Hence we find an edge state solution that is predominantly localized at a single edge, as shown in Fig. 6.7(b). This exponential suppression of one sublattice (in this case sublattice A) and the fact that the functional form of the edge state solution does not change with $M \neq 0$ is most obvious in the inset of Fig. 6.7(b).

Hence, we find that a staggered potential localizes the edge state towards a single edge in a zigzag graphene nanoribbon. In this respect, the behavior of the edge state with a staggered potential is much more intuitive than without. At first glance such

a staggered potential seems rather artificial, but we will discuss in Chapter 7 how antiferromagnetic ordering naturally induces such a staggered potential.

6.3.4 Effects of next-nearest neighbor hopping

Finally, we turn to the effects of next-nearest neighbor hopping and start by considering the half-infinite graphene sheet with a single zigzag edge.

For this case, Eqs. (6.19) and (6.20) can be extended to include next-nearest neighbor hopping:

$$\begin{aligned} E \phi_A(j) = & -t \phi_B(j) + t e^{-\gamma} \phi_B(j-1) \\ & + t' e^{-\gamma} \phi_A(j+1) + t' e^{-\gamma} \phi_A(j-1) + t'(2 - e^{-2\gamma}) \phi_A(j), \end{aligned} \quad (6.45)$$

$$\begin{aligned} E \phi_B(j) = & -t \phi_A(j) + t e^{-\gamma} \phi_A(j+1) \\ & + t' e^{-\gamma} \phi_B(j+1) + t' e^{-\gamma} \phi_B(j-1) + t'(2 - e^{-2\gamma}) \phi_B(j), \end{aligned} \quad (6.46)$$

where we used the identity $2 - e^{-2\gamma} = -2 \cos(ka)$. At the boundary, these equations reduce to

$$\begin{aligned} E \phi_A(1) = & -t \phi_B(1) + t e^{-\gamma} \phi_B(0) \\ & + t' e^{-\gamma} \phi_A(2) + t'(2 - e^{-2\gamma}) \phi_A(1), \end{aligned} \quad (6.47)$$

and

$$E \phi_B(0) = t e^{-\gamma} \phi_A(1) + t' e^{-\gamma} \phi_B(1) + t'(2 - e^{-2\gamma}) \phi_B(0). \quad (6.48)$$

Unlike for the case of nearest-neighbor hopping, an analytical solution of these equations does not seem feasible. Thus, we will use perturbation theory to study the effects of next-nearest neighbor hopping, as in Ref. [183]. To this end, note that the edge state solution for $t' = 0$ (Eq. (6.24)),

$$\phi_A(j) = 0, \quad \phi_B(j) = \sqrt{1 - e^{-2\gamma}} e^{-\gamma j}, \quad (6.49)$$

is *almost* a solution for $t' \neq 0$, too. In particular, it solves Eqs. (6.45),(6.46) and (6.47) for $E = 3t'$.⁷ However, it solves Eq. (6.48) only for $E = 2t'$, and thus is not an eigenstate. Therefore, apart from a trivial overall energy shift, the next-nearest neighbor hopping only influences the edge state close to the graphene edge. In fact, the previous analysis shows that for the edge state, next-nearest neighbor hopping corresponds to putting an additional on-site potential $-t'$ on the edge atoms. However, since the edge state is localized at the edge, it is then influenced significantly by $t' \neq 0$.

⁷ Note in this respect that the next-nearest neighbor hopping also shifts the bulk energy spectrum by $3t'$, as discussed in Section G.2.2.

In first order perturbation theory, the energy of the edge state is shifted by

$$\begin{aligned}\Delta E(k) &= \langle \varphi | \Delta H | \varphi \rangle / \langle \varphi | \varphi \rangle \\ &= \left(2t' + 3t' \sum_{j=1}^{\infty} e^{-2\gamma j} \right) (1 - e^{-2\gamma}) \\ &= \left(-t' + 3t' \frac{1}{1 - e^{-2\gamma}} \right) = 4t' + 2t' \cos(ka).\end{aligned}\quad (6.50)$$

Thus, the edge state energy is shifted by the bulk value $3t'$ at the K -points, but only by $2t'$ at $k = \pm \frac{\pi}{a}$ where the edge state is maximally localized. The previously perfectly flat edge state band hence acquires a dispersion given by $\Delta E(k)$.

Close to the K -points, the energy shift $\Delta E(k)$ can be expanded as

$$\Delta E(k) = 3t' \pm \sqrt{3}at' \left(k \pm \frac{2\pi}{3a} \right) + \mathcal{O} \left(\left(k \pm \frac{2\pi}{3a} \right)^2 \right), \quad (6.51)$$

i.e. the next-nearest neighbor hopping introduces corrections to the edge state energy dispersion linear in $q = k \pm \frac{2\pi}{3a}$. Note that we could neglect t' for bulk states in the low-energy regime, since next-nearest neighbor hopping only introduced terms quadratic in \mathbf{q} , as discussed in Appendix G. However, since t' enters the edge state dispersion already to linear order in q , *next-nearest neighbor hopping may not be neglected in this case*.

We do not attempt to do perturbation theory for the zigzag nanoribbon, but we can extrapolate the results for a single edge to the case of two edges in a nanoribbon. Hence, we expect that the almost dispersionless bands, that we obtained in nearest-neighbor tight-binding, will also obtain a finite dispersion for $t' \neq 0$. This dispersion will compete with the energy splitting due to the finite width of the ribbon, Eq. (6.39). However, since this splitting decreases exponentially with the nanoribbon width, the width-independent dispersion from next-nearest neighbor hopping will usually dominate.

We demonstrate these predictions in Fig. 6.8(a), where we show the band structure of a zigzag nanoribbon in next-nearest neighbor tight-binding approximation. Note that we have subtracted the trivial energy shift $3t'$ here, and we will use this convention in the remainder of this work, when considering next-nearest neighbor hopping. In the band structure, we observe two edge state bands (arising from the two edges of the nanoribbon) that now have a significant dispersion, in contrast to the almost flat bands for nearest-neighbor hopping only (see Fig. 6.6(a)).

In Fig. 6.8(b) we show the probability density of the edge state across the nanoribbon for a given value of k . Since the ribbon has a reflection symmetry between the two edges, the Bloch eigenstate for a fixed value of the wave vector k still occupies both edges with equal probability. However, due to the dispersion of the edge state band, there is an energy range where in addition to the “bonding state” we also have an “anti-bonding” state at the same energy (and, additionally, an extended state). Thus,

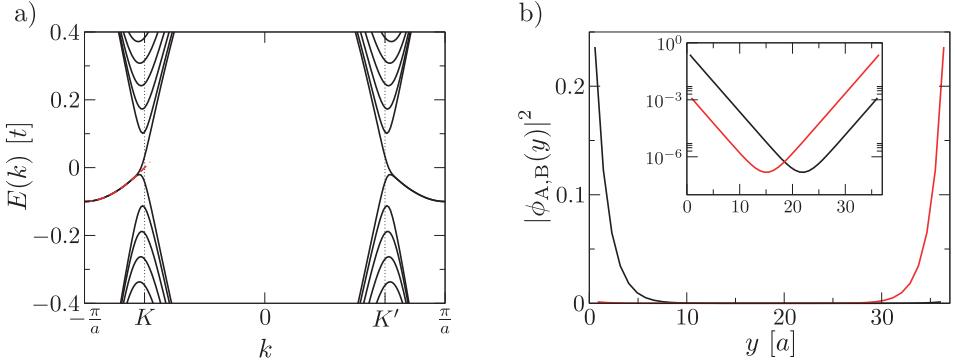


Figure 6.8 – A zigzag graphene nanoribbon in next-nearest neighbor tight-binding approximation for $t' = 0.1t$. (a) shows the band structure of the ribbon, where we have subtracted the trivial energy shift $3t'$. For comparison, the red dashed curve gives the energy dispersion of the edge state obtained from perturbation theory, Eq. (6.50). In (b) we show the probability density profile of the edge state across the nanoribbon, for $k = -2.4/a$. The inset shows the same density profile on a logarithmic scale. Again, the probability density for sublattice A is shown in red, for sublattice B in black. In both (a) and (b) the nanoribbon had a width $W = 64a/\sqrt{3}a$.

we can form linear combinations of these states to obtain eigenstates that are located at a single edge⁸. In addition to that, the next-nearest neighbor hopping also changes the character of the wave function, as seen in the inset of Fig. 6.8(b): Whereas the edge state resided on a single sublattice close to the boundary in the examples of the previous section, $t' \neq 0$ leads to an increased (but still small) occupation probability on the other sublattice. This has important consequences, that are discussed in the next section.

We conclude the characterization of the graphene edge state by discussing how to include the effects of next-nearest neighbor hopping in the effective Hamiltonian. Note that we have argued that the influence of the next-nearest neighbor hopping on the edge state is equivalent to a potential $-t'$ at the zigzag edge atoms. In the effective Hamiltonian we can include such a potential as

$$V_{t'}(y) = 3t' - \frac{\sqrt{3}a}{2}t'\delta(y - y_{\text{edge}}), \quad (6.52)$$

⁸ Note that the “anti-bonding” state has a slightly different wave vector than the “bonding state”, and thus a slightly different shape of the wave function. This difference in the wave vectors is due to the energy splitting because of the finite width of the nanoribbon. However, since this energy splitting decreases exponentially with the ribbon width, so does the difference in the wave vector. Thus, the linear combination of “bonding” and “anti-bonding” state is to a good approximation localized at a single edge only.

where y_{edge} is the y -coordinate of the boundary. Note that $\sqrt{3}a/2$ is the length of the graphene unit cell in y -direction. Such a potential will influence strongly a state localized at the boundary, but only little a state extended through the whole system. For a single zigzag edge, we obtain the energy shift in first order perturbation theory within the effective Hamiltonian as

$$\Delta E(q) = -2\tau q \int_0^{\infty} e^{-2\tau q y} V_t(y) = 3t' - \sqrt{3}at'\tau q, \quad (6.53)$$

in agreement with Eq. (6.51).

6.4 Edge state transport in zigzag nanoribbons

6.4.1 Model-dependence of edge state transport—some peculiar examples

As mentioned in the introduction, numerical studies of edge state transport in zigzag nanoribbons have, up to now, only considered graphene in nearest-neighbor tight-binding approximation [177–180]. The main finding of these accounts is that edge state transport is influenced only little by edge defects. This is a rather counterintuitive result: Since the probability density of the edge state is strongly localized to the edges of the ribbon, one would naively expect it to scatter strongly from defects at the edge. In fact, edge state transport holds even more surprises, as we will demonstrate on illustrative examples in this section.

For this, we use the numerical framework introduced in the first part of this work to calculate transmission probabilities and density as well as current profiles of zigzag nanoribbons. However, before presenting the calculations, it is worthwhile to spare some thoughts about general aspects of numerical studies on edge state transport.

Usually, numerical transport studies of nanoribbons compute the transmission probability as a function of Fermi energy E_F . The range of energies considered is commonly on the order of t , implying that the almost flat edge state band in nearest-neighbor approximation only contributes a few—or often only a single—point in the numerically computed transmission spectrum. Computing the transmission as a function of Fermi energy is based on the idea of mimicking the effect of a gate voltage. In an ordinary two-dimensional electron gas in a semiconductor heterostructure, the gate voltage V_{gate} and the Fermi energy E_F are in fact linearly proportional, since a capacitively coupled gate induces a charge $Q = CV_{\text{gate}}$ (C is the capacitance of the structure) and the density of states is a constant. This is not the case anymore in bulk graphene, where the density of states is linear in energy. Even worse in nanoribbons, the edge states in nearest-neighbor tight-binding have a very high density of states due to the almost flat edge state band. Hence, changing the Fermi energy linearly will grossly underestimate the gate voltage region where the edge states play a dominant role in transport. Note that this problem is less severe, if the edge states acquire a dispersion, such as due

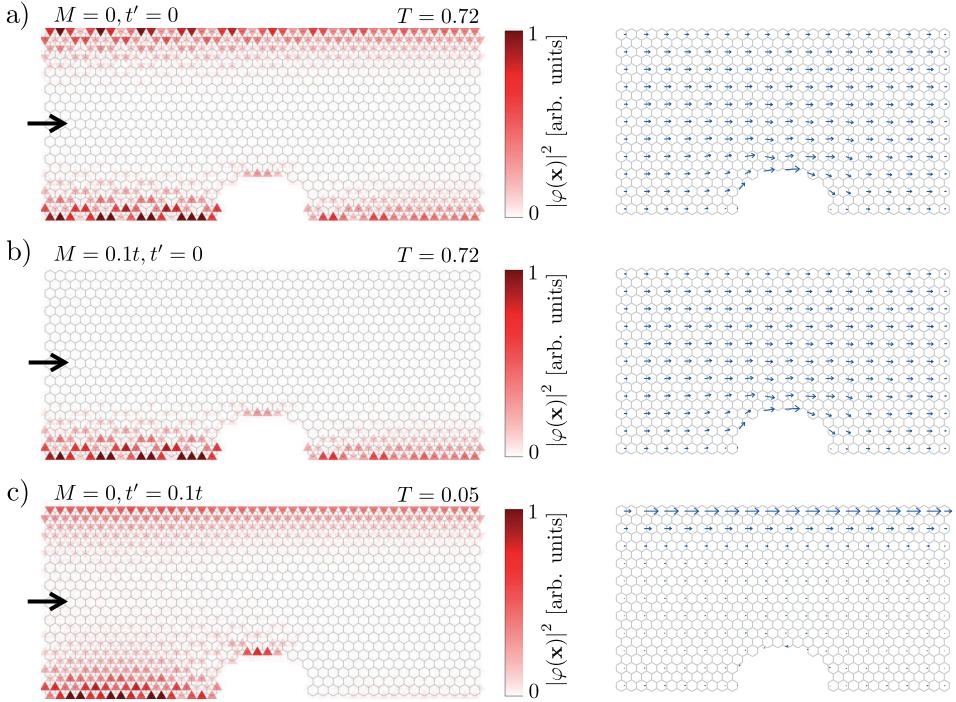


Figure 6.9 – Density profile (left panels) and current density profile (right panels) of the edge state ($k = -2.4/a$) in a zigzag nanoribbon (width $W = 34a/\sqrt{3}$) with a large obstacle on the lower edge. (a) shows the edge state in nearest-neighbor approximation ($t' = 0$ and $M = 0$), (b) for a finite staggered potential ($t' = 0$ and $M = 0.1t$), and (c) in next-nearest neighbor approximation ($t' = 0.1t$, $M = 0$). In all examples, the channels are incoming from the left.

to next-nearest neighbor hopping or other perturbations. Furthermore, we have seen in the previous sections that the extent of the edge state wave function is determined by the Bloch vector k , and is largely independent of the particular model employed to describe graphene. Thus, instead of Fermi energy, we will use k in order to characterize the edge state in this and the following sections.

In order to demonstrate the peculiarities of edge state transport, we consider a zigzag nanoribbon with a rather large edge defect, as shown in Fig. 6.9. The wave vector $k = -2.4/a$ of the edge state was chosen such that the extent of the edge state in the perpendicular direction is smaller than the size of the defect.

We begin by considering the case of graphene in nearest-neighbor approximation. Even though the extent of the probability density at one of the edges is smaller than

the defect itself, as shown in Fig. 6.9(a), the wave function also comprises the density at the opposite edge, and in this sense extends through the whole width of the ribbon. Since the edge state in this situation is situated on both edges at the same time, it is hard to predict *a priori* the effect of a disturbance on a single edge only. Still, intuition would suggest that an edge defect should result in strong scattering. Yet, we find a rather large transmission probability of $T = 0.72$. Furthermore, if we made the defect smaller, we would find almost perfect transmission, if only a single lattice site was removed [177–179].

In order to explain this seeming contradiction between a state localized at the edge and its insensitiveness to edge disorder, Refs. [177] and [179] considered the current density of the edge state. Here, we will follow their arguments using the effective Hamiltonian of graphene.

Within the effective Dirac Hamiltonian, the velocity operator is given as

$$\mathbf{v} = v_F \begin{pmatrix} \tau \sigma^x \\ \sigma^y \end{pmatrix}. \quad (6.54)$$

Hence, we find the current density of a state φ as

$$\mathbf{j}(\mathbf{x}) = e\varphi^\dagger(\mathbf{x})\mathbf{v}\varphi(\mathbf{x}). \quad (6.55)$$

From the edge state solution (6.35) we find the current density in ribbon direction as

$$\begin{aligned} j_x(y) &\cong \phi_A(y)\phi_B(y) \\ &= \frac{1}{2} \left(\cosh(zW) - \cosh(2z(W/2 - y)) \right). \end{aligned} \quad (6.56)$$

In Fig. 6.10 we show both the probability and current density profiles of the edge state. Note that the current densities have been normalized such that the integrated current carries one quantum of conductance⁹. The current density always has a maximum in the middle of the ribbon, whereas it vanishes at the edges, where the wave function of one of the sublattices must be zero, as seen from the zigzag boundary conditions (6.34). When the edge state decays slowly and extends far into the ribbon, the current density profile is nearly sinusoidal (more precisely, it is parabolic for $z = 0$), and is almost constant throughout the whole ribbon, if the edge state is strongly localized to the edge.

This behavior may be understood from the lattice structure of graphene: In nearest neighbor tight-binding approximation, current flow requires hopping from one sublattice to the other. Since the edge state close to a single edge is localized on a single

⁹ The rationale behind this normalization is that in ballistic transport every channel contributes e^2/h to the conductance: The current density depends on the velocity of the channel, but this dependence is cancelled exactly by the density of states. In this normalization, the current density $j_x(y)$ of the state becomes essentially the energy-dependent current density $j_x(y, E)$ introduced in Appendix A.

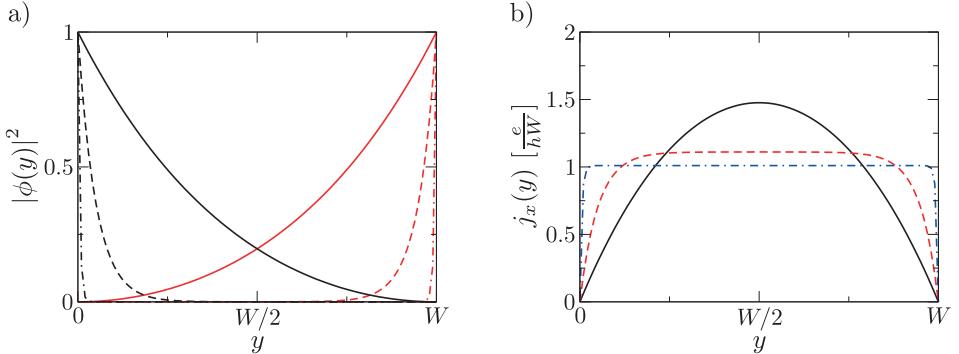


Figure 6.10 – (a) Probability density and (b) current profile of the edge state as a function of the perpendicular position y in a zigzag nanoribbon of width W . (a) and (b) show the behavior for different decay constants of the edge state: $zW = 1$ (solid lines), $zW = 10$ (dashed lines), and $zW = 100$ (dash-dotted lines). In (a), sublattice A (B) is encoded in red (black), and the wave functions have been normalized to unity at the edges.

sublattice only, current is suppressed at the edges, and reaches a maximum in the middle, where the tails of the exponential decay overlap¹⁰. The fact that current is flowing mainly in the middle of the ribbon explains the fact that edge defects have only little influence of the edge state. As seen in Fig. 6.9(a) the current density merely meanders around the obstacle. Furthermore, the value of the transmission, $T \approx 3/4$ can be understood from the ratio of the extent of the obstacle and the ribbon width which is around 1/4. In fact, we would still find a sizeable transmission, if we introduced defects at *both* edges.

Thus, although the edge state density is strongly localized at the edges, its current density rather behaves as expected from a bulk state. Note that the current density is zero for all states at the edges of a zigzag nanoribbon in nearest-neighbor approximation, since this is dictated by the boundary conditions (6.34). Thus, we also find extended states, where the density profile differs from the current profile. However, the difference is most drastic for the edge state, where the density at the edges is exponentially larger than in the middle.

We next turn to the case of a finite staggered potential $M = 0.1t$, depicted in Fig. 6.9(b). In contrast to the previously considered case $M = 0$, the edge state is now localized to a single edge only. Here, we have chosen the edge state at $E \approx -M$ which is localized at the same edge as the defect. Even more than in the previous case, intuition

¹⁰The maximum of the current density is in the middle, where the contributions from both sublattices are already very small, due to the exponential decay. The resulting velocity (and also the total current) of the state is thus exponentially small, which is reflected in the almost dispersionless edge state band.

would suggest that the state should be scattered back with a large probability. Yet, we find that the channel is transmitted again with a very high transmission probability $T = 0.72$, which is identical to the case $M = 0$. This can be understood from the shape of the wave function (6.41), which is, up to different weights for sublattice A and B, identical to the wave function for $M = 0$, Eq. (6.35). Thus, the current density profiles are equal in both cases, as can be seen in Fig. 6.9(b). Even the pattern of the reflected wave is identical in both situations. In fact, in this situation the edge state would be scattered as much from a defect on the *opposite* edge. Hence, we find the seemingly paradoxical situation that a state localized at a single edge, and thus being most asymmetric with regard to both edges, features a perfectly symmetric current density, and thus is affected equally by defects on either edge.

Finally, we consider the case of finite next-nearest neighbor hopping $t' = 0.1t$, which also allows for edge state solutions located at a single edge, just as $M \neq 0$. In this situation we find that the edge state located at the defective edge is scattered back almost perfectly ($T = 0.05$), whereas the state on the opposite edge can pass unaffected (see Fig. 6.9(c)). In addition, the current density is mainly located at the edge, in accordance with the probability density. Thus, the transport behavior of the edge state for $t' \neq 0$ differs fundamentally from the previous examples, although the probability densities look much alike.

As discussed above, the edge state in a ribbon with nearest-neighbor hopping only is comprised of contributions from both edges, because of the particle-hole symmetry associated with \mathcal{P}_z . Because of this symmetry, both edges are inevitably coupled. From this point of view, the counter-intuitive transport properties of the edge state, namely the insensitiveness to edge disorder, may not seem too surprising. Yet, even after breaking this particle symmetry with a mass term, localizing the edge state to a single edge only, the transport properties of the edge state remain unchanged. In contrast, next-nearest neighbor hopping, which also breaks \mathcal{P}_z , changes the transport properties fundamentally. It is thus not enough to break the particle-hole symmetry \mathcal{P}_z alone to change edge state transport. In the examples of this section, the shape of the current density was most useful to understand the peculiar transport features. Thus, in the next section, we will examine the shape of the current density systematically.

6.4.2 Where does the current flow?

In order to investigate the behavior of the current density profile of the edge state in dependence of additional terms to the graphene Hamiltonian in nearest-neighbor tight-binding approximation, we consider the following models of a zigzag nanoribbon:

- I. nearest-neighbor hopping only.
- II. nearest-neighbor hopping with a constant staggered potential M .
- III. finite next-nearest neighbor hopping t' ,

IV. nearest-neighbor hopping with a position-dependent staggered potential $M(\mathbf{x})$.

$$M(\mathbf{x}) = \begin{cases} M_0 e^{-\bar{\gamma}y} & \text{for } y < W/2, \\ M_0 e^{-\bar{\gamma}(W-y)} & \text{for } y > W/2. \end{cases} \quad (6.57)$$

Thus, the magnitude of the staggered potential decays exponentially away from the edges. Such a model may be appropriate for describing edge magnetism (see Chapter 7).

V. nearest-neighbor hopping with a position-dependent potential $V(\mathbf{x})$:

$$V(\mathbf{x}) = V_0 \left(\frac{1}{2} - \frac{y}{W} \right). \quad (6.58)$$

This is a potential that changes linear from $-V_0/2$ at one edge to $V_0/2$ at the other edge. This may model the effects of side gates (see Ref. [176]).

Note that models II–V break the particle-hole symmetry associated with \mathcal{P}_z and thus allow the localization of the edge state on a single edge only.

In Fig. 6.11 we show a comparative overview of the band structure and the edge state probability as well as current density profile for all five models (for a discussion of the respective observables, see Appendix A).

Fig. 6.11(a) and (b) show a almost constant current density profile with a maximum in the middle, as discussed in the previous section, whereas in (c)–(e), the current density profile has distinct maxima at the edges, in accordance with the probability density. In these calculations, we chose the Fermi energy such that we have an edge state with a fixed value of $k = -2.4/a$. Due to the band-bending in models III–V there are three channels at the Fermi energy in Fig. 6.11(c) (two edge states and one extended state) and two channels in (d) and (e) (one edge state and one extended state). The finite current density in the middle of the nanoribbon in (c)–(e) is due to the extended state. In contrast to the current density that differs fundamentally between the different models, the probability density of the edge state wave function is almost identical for all models—only on a logarithmic scale it is possible to observe differences.

When we observe a difference in the current density profile for the various models, this can be either due to a change in the wave function, or due to a change in the current density operator. The current density operator is identical for models I, II, IV and V, since all these comprise only nearest-neighbor hopping¹¹. The difference in these models thus must be due to changes in the wave function. In contrast, next-nearest neighbor hopping (model III) also introduces changes to the current density operator. However, it turns out that also in this case the change in the wave function is the dominant source for altering the current density profile. This can be seen by

¹¹The current density operator in tight-binding only depends on hopping elements, not on the on-site energies (see Appendix A).

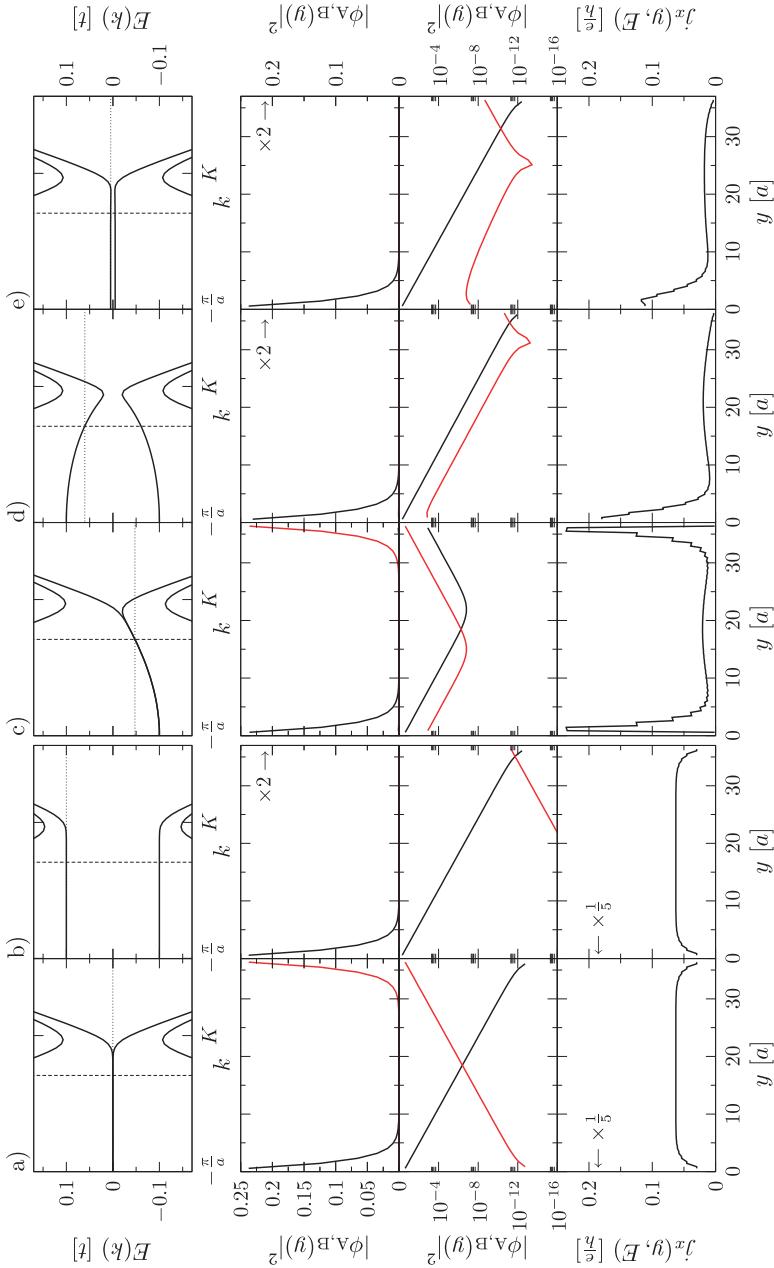


Figure 6.11 – Overview over the edge state behavior for different models of a zigzag nanoribbon (width $W = 64a/\sqrt{3}$): (a) nearest neighbor hopping only (model I), (b) a constant staggered potential with $M = -0.1t$ (model II), (c) next-nearest neighbor hopping $t' = 0.1t$ (model III), (d) a position-dependent staggered potential $M(\mathbf{x})$ with $M_0 = -0.1t$ and $\overline{\gamma} = 1/a$ (model IV), and (e) a position-dependent potential $V(\mathbf{x})$ with $V_0 = 0.01t$ (model V). The upper panels show the band structure in the vicinity of the edge state. We consider edge states with wave vector $k = -2.4/a$; in the plot of the band structure, this k -value is indicated as a dashed line and the corresponding Fermi energy as a dotted line. The second and third row of panels show the probability density profile of the edge state on a linear and logarithmic scale, respectively, whereas the lower panels show the current density profiles for all states at the Fermi energy.

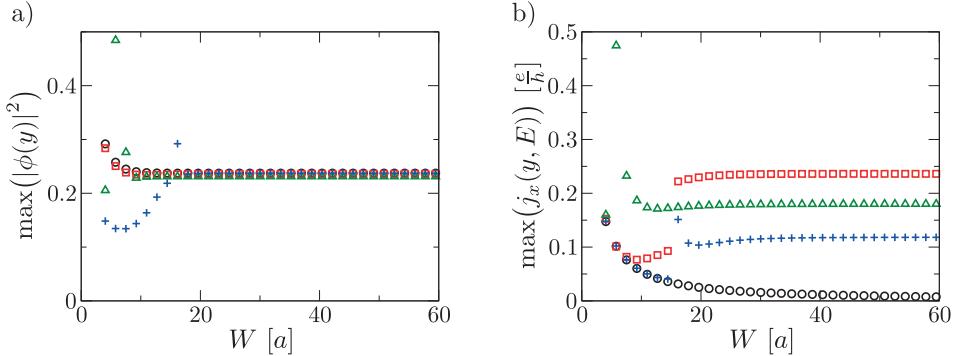


Figure 6.12 – Maximum value of the (a) probability and (b) current density of the edge state ($k = -2.4/a$) as a function of the ribbon width W , for nearest-neighbor hopping only (\circ), next-nearest neighbor hopping (\square), a position-dependent staggered potential (\triangle), and a position-dependent potential (+). Note that we have divided the probability density of the models IV and V by 2, since there the edge state is localized on a single edge only.

comparing the current densities computed on the one hand with a current operator including t' and, on the other hand, with a current operator where t' is set to zero artificially.

In Fig. 6.11(a) and (b) we observe that the occupation probability of a sublattice decreases monotonically from one edge to the other. This is not the case for (c)–(e): There, the probability density increases again towards the opposite edge, so that for every edge state that occupied a single sublattice only, there is an enhanced occupation probability of the opposite sublattice. Furthermore, the probability densities show a similar functional behavior in the vicinity of the edges, and seem in a sense “linked” to each other.

Before pursuing these observations further, we briefly revisit the statement of the previous section that the current density of the edge state in nearest-neighbor approximation behaves like a bulk state. A bulk state extends through the whole system. In particular this means that the occupation probability for a single lattice site scales inversely proportional with the width of the nanoribbon. In contrast, a true edge state remains unchanged when the width of the nanoribbon is varied. In Fig. 6.12(a) we show the maximum value of the probability density for the models I and III–V (As discussed in the previous section, model II behaves just like I in this regard, and is thus not shown explicitly). Except for small widths, where finite size effects play a role, all models show a constant maximum of the probability density, independent of the ribbon width. This is the signature of an edge state. In contrast, only models III–V show a constant maximum of the current density in Fig. 6.12(b) whereas the maximum of the current density in model I decreases as the inverse of the nanoribbon width. Thus,

again the current density of the edge state in nearest-neighbor tight-binding approximation shows bulk-state-like behavior, whereas models III–V behave like a true edge state in every sense.

The enhanced occupation probability of the opposite sublattice and a current flowing at the edges are necessarily correlated: In order to obtain current flowing close to the edge, the value of the wave function on the sublattice that was close to zero before must be increased, as the current density is given by the product of the wave functions on both sublattices, Eq. (6.56). Furthermore, it seems that a maximum in the current density at the edge is always linked to a finite bending of the edge state band, in particular a bending beyond the exponentially small energy splitting due to the finite width. In the next section, we will further elaborate on these connections.

6.4.3 Effects of perturbations on the current flow

In order to better understand, under which circumstances the current density changes from bulk-like to edge-like, we study the corrections to the edge state of a single edge in perturbation theory within the effective Dirac Hamiltonian. The results will then be interpolated to the case of a nanoribbon.

To this end we consider again a half-infinite graphene sheet with a single zigzag edge as in Fig. 6.2(b), governed by the Dirac Hamiltonian H_τ , Eq. (6.6), including a perturbation:

$$H = H_\tau + V^\mu, \quad (6.59)$$

where

$$V^\mu = g(\partial_x, \partial_y, y) \sigma^\mu. \quad (6.60)$$

Here, $\mu = 0, x, y, z$, and $g(\partial_x, \partial_y, y)$ is a function that may contain arbitrary derivatives, but only depends on the coordinate y perpendicular to the zigzag edge. This perturbation potential includes examples such as a magnetic vector potential ($V^\mu \sim \sigma^x, \sigma^y$), a staggered potential ($V^\mu \sim \sigma^z$), a potential or next-nearest neighbor hopping ($V^\mu \sim \sigma^0$). The system then still exhibits translational symmetry in the x -direction, and the solution has the Bloch form (6.28). As a consequence, Eq. (6.59) turns into a one-dimensional differential equation for $\phi(y)$:

$$(H_\tau(q) + V^\mu(q)) \phi(y) = E \phi(y), \quad (6.61)$$

where

$$H_\tau(q) = \hbar v_F (\tau q \sigma^x + (-i \partial_y) \sigma^y) \quad (6.62)$$

is the unperturbed Dirac Hamiltonian and

$$V^\mu(q) = g(iq, \partial_y, y) \sigma^\mu \quad (6.63)$$

the perturbation for a Bloch vector q . Within this subspace, we can then apply non-degenerate perturbation theory (for example, see Ref. [153, Chap. 11]).

The unperturbed Hamiltonian $H_0(q)$ together with the zigzag boundary condition $\phi_A(0) = 0$ has the already known edge state solution, Eq. (6.24),

$$\phi^{\text{edge}}(y) = \sqrt{-2\tau q} e^{\tau q y} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{with energy } E^{\text{edge}} = 0 \quad (6.64)$$

as well as the bulk solutions

$$\phi^\pm(z, y) = \frac{1}{\sqrt{\pi}} \begin{pmatrix} \sin(zy) \\ \pm \sin(zy + \theta) \end{pmatrix} \quad \text{with energy } E^\pm(q, z) = \pm \hbar v_F \sqrt{q^2 + z^2}, \quad (6.65)$$

where the angle θ is determined by the equation

$$e^{i\theta} = \frac{\tau q + iz}{\sqrt{q^2 + z^2}} = \frac{\sqrt{q^2 + z^2}}{\tau q - iz}, \quad (6.66)$$

and z is a positive real number. The eigenstates have been normalized such that the completeness relation

$$\phi^{\text{edge}}(y) (\phi^{\text{edge}}(y'))^\dagger + \sum_{\tau=\pm} \int_0^\infty dz \phi^\tau(z, y) (\phi^\tau(z, y'))^\dagger = \mathbb{1} \delta(y - y') \quad (6.67)$$

holds.

Then, the energy shift $\Delta E(q)$ in first order perturbation theory is given as

$$\begin{aligned} \Delta E(q) &= \langle \phi^{\text{edge}} | V^\mu | \phi^{\text{edge}} \rangle \\ &= -2\tau q \int_0^\infty dy' e^{\tau q y'} \begin{pmatrix} 0 \\ 1 \end{pmatrix}^\dagger \sigma^\mu \begin{pmatrix} 0 \\ 1 \end{pmatrix} g(iq, \partial_{y'}, y') e^{\tau q y'}. \end{aligned} \quad (6.68)$$

The expression for $\Delta E(q)$ contains an expectation value of a Pauli matrix σ^μ with an eigenstate of σ^z . Hence, we find that the energy shift $\Delta E(q) = 0$ for $V^\mu \sim \sigma^x, \sigma^y$. This is not surprising, as such a perturbation does not change the off-diagonal form of the Hamiltonian. Thus, the differential equations for the two sublattices still decouple for $E = 0$ and there will be a zero-energy solution. On the contrary, if $V^\mu \sim \sigma^0, \sigma^z$, we find a finite (q -dependent) energy shift

$$\Delta E(q) = \mp 2\tau q \int_0^\infty dy' e^{2\tau q y'} g(iq, q, y'), \quad (6.69)$$

where the upper sign refers to $V^\mu \sim \sigma^0$ and the lower sign to $V^\mu \sim \sigma^z$. Thus, any perturbation $V^\mu \sim \sigma^0, \sigma^z$ will in general lead to band bending.

The first-order correction to the wave function is given as

$$\Delta \phi(y) = \sum_{\tau=\pm} \int_0^\infty dz \frac{\langle \phi^\tau(z) | V^\mu | \phi^{\text{edge}} \rangle}{0 - E^\tau(q, z)} \phi^\tau(z, y) \quad (6.70)$$

which is evaluated to

$$\Delta\phi_A(y) = \frac{1}{\sqrt{\pi}} \int_0^\infty dz \frac{1}{E^+(q, z)} (\langle \phi^-(z) | V^\mu | \phi^{\text{edge}} \rangle - \langle \phi^+(z) | V^\mu | \phi^{\text{edge}} \rangle) \sin(zy) , \quad (6.71\text{a})$$

$$\Delta\phi_B(y) = -\frac{1}{\sqrt{\pi}} \int_0^\infty dz \frac{1}{E^+(q, z)} (\langle \phi^+(z) | V^\mu | \phi^{\text{edge}} \rangle + \langle \phi^-(z) | V^\mu | \phi^{\text{edge}} \rangle) \sin(zy + \theta) , \quad (6.71\text{b})$$

where we made use of the relation $E^+(q, z) = -E^-(q, z)$. These equations can be simplified further by the observation that

$$\begin{aligned} \langle \phi^+(z) | V^\mu | \phi^{\text{edge}} \rangle &= \sqrt{\frac{-2\tau q}{\pi}} \int_0^\infty dy' \begin{pmatrix} \sin(zy') \\ \sin(zy' + \theta) \end{pmatrix}^\dagger \sigma^\mu \begin{pmatrix} 0 \\ 1 \end{pmatrix} g(iq, \partial_{y'}, y') e^{-\tau q y'} \\ &= \begin{cases} +\langle \phi^-(z) | V^\mu | \phi^{\text{edge}} \rangle & \text{for } V^\mu \sim \sigma^x, \sigma^y, \\ -\langle \phi^-(z) | V^\mu | \phi^{\text{edge}} \rangle & \text{for } V^\mu \sim \sigma^0, \sigma^z. \end{cases} \end{aligned} \quad (6.72)$$

Inserting these expressions into Eqs. (6.71) we find that $\Delta\phi_A(y) = 0$ and $\Delta\phi_B(y) \neq 0$ for $V^\mu \sim \sigma^x, \sigma^y$, i.e. the edge state remains localized on a single sublattice. Again, this is not surprising, since σ^x and σ^y do not change the off-diagonal structure of the Dirac Hamiltonian, and the corresponding zero-energy solutions are located on a single sublattice only. In particular, this applies to the case of a weak magnetic field, where the magnetic vector potential creates a perturbation $\sim \sigma^x, \sigma^y$.

In contrast, if $V^\mu \sim \sigma^0, \sigma^z$, we find

$$\Delta\phi_A(y) = -\frac{2}{\sqrt{\pi}} \int_0^\infty dz \frac{1}{E^+(q, z)} \langle \phi^+(z) | V^\mu | \phi^{\text{edge}} \rangle \sin(zy + \theta) , \quad (6.73\text{a})$$

$$\Delta\phi_B(y) = 0 . \quad (6.73\text{b})$$

Thus, terms proportional to σ^0 and σ^z will in general introduce a finite occupation probability on the sublattice that was strictly zero in the unperturbed Hamiltonian.

We now turn to evaluating Eq. (6.73a):

$$\begin{aligned} \Delta\phi_A(y) &= -\frac{2}{\pi} \sqrt{-2\tau q} \int_0^\infty dz \int_0^\infty dy' \frac{1}{\hbar v_F \sqrt{q^2 + z^2}} \\ &\quad \begin{pmatrix} \sin(zy') \\ \sin(zy' + \theta) \end{pmatrix}^\dagger \sigma^{0,z} \begin{pmatrix} 0 \\ 1 \end{pmatrix} g(iq, \partial_{y'}, y') e^{-\tau q y'} \sin(zy) \\ &= \pm \frac{\sqrt{-2\tau q}}{2\pi \hbar v_F} \int_0^\infty dy' g(iq, q, y') e^{\tau q y'} \int_{-\infty}^\infty dz \left(\frac{e^{iz(y'+y)}}{\tau q - iz} - \frac{e^{-iz(y'-y)}}{\tau q + iz} \right) , \end{aligned} \quad (6.74)$$

where the upper sign refers to σ^0 and the lower sign to σ^z . In order to arrive at this expression, we used $\sin(x) = \frac{1}{2i}(e^{ix} - e^{-ix})$ and Eq. (6.66). We can simplify this

expression further by making use of the identities

$$\int_{-\infty}^{\infty} dz e^{-izy} \frac{1}{\tau q + iz} = \int_{-\infty}^{\infty} dz e^{izy} \frac{1}{\tau q - iz} = -2\pi \Theta(y) e^{\tau q y}, \quad (6.75)$$

where $\Theta(y)$ is the Heaviside step function. Eq. (6.75) can be easily obtained by contour integration. We finally obtain

$$\begin{aligned} \Delta\phi_A(y) = & \mp \frac{\sqrt{-2\tau q}}{\hbar v_F} \left(e^{\tau q y} \int_0^{\infty} dy' g(iq, q, y') e^{2\tau q y'} \right. \\ & \left. - e^{-\tau q y} \int_y^{\infty} dy' g(iq, q, y') e^{2\tau q y'} \right), \end{aligned} \quad (6.76)$$

where again the upper sign is for $V^\mu \sim \sigma^0$ and the lower sign for $V^\mu \sim \sigma^z$.

In order to get a feel for the nature of the perturbed wave function we now evaluate Eq. (6.76) for the models of the previous section:

- $V^\mu = V_0 \sigma^0$ or $V^\mu = M \sigma^z$: For a constant potential or staggered potential we obtain

$$\Delta\phi_A(y) = 0. \quad (6.77)$$

This is obvious for the constant potential, as this only leads to an overall energy shift. For the staggered potential this is in agreement with the findings of Section 6.3.3, where the occupation probability on sublattice A originated solely due to the second edge in the ribbon.

- $t' \neq 0$, $V^\mu = V_t(y) \sigma^0 = 3t' - \frac{\sqrt{3}a}{2} t' \delta(y - y_{\text{edge}})$: For finite next-nearest neighbor hopping we obtain

$$\Delta\phi_A(y) = \frac{t'}{t} \sqrt{-2\tau q} e^{\tau q y} = \frac{t'}{t} \phi_B(y). \quad (6.78)$$

- $V^\mu = M_0 e^{-\bar{\gamma}y} \sigma^z$: For a position-dependent staggered potential we obtain

$$\Delta\phi_A(y) = \frac{M (e^{-\bar{\gamma}y} - 1)}{\hbar v_F (2\tau q - \bar{\gamma})} \phi_B(y). \quad (6.79)$$

- $V^\mu = V_0/W (W/2 - y) \sigma^0$: For a position-dependent potential we obtain

$$\Delta\phi_A(y) = \frac{V_0 y}{\hbar v_F W (-2\tau q)} \phi_B(y). \quad (6.80)$$

Hence, we find that the wave function of sublattice A always is of the form $f(y) \phi_B(y)$, where $f(y)$ is some function of y . In fact, this form is rather general, as can be seen from Eq. (6.76): Provided that the perturbing function $g(iq, q, y)$ is sufficiently benign—and only for such functions perturbation theory is applicable—partial integration of the second term in Eq. (6.76) will always lead to terms proportional to $\phi_B(y) = \sqrt{-2\tau q} e^{\tau q y}$.

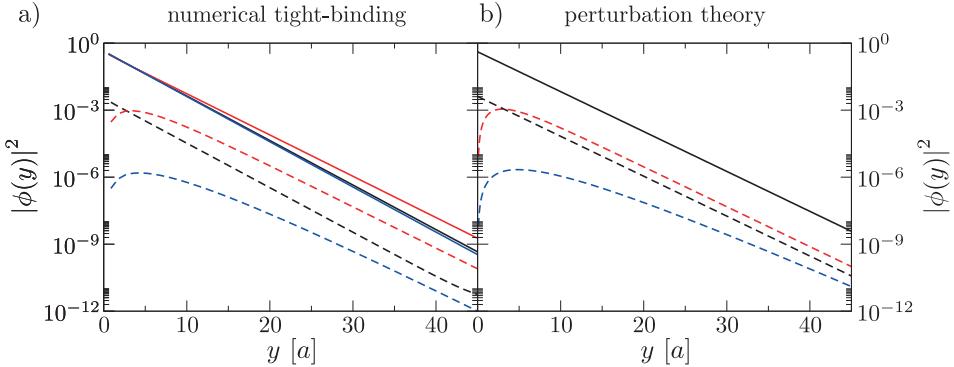


Figure 6.13 – Comparison of edge state probability densities for (a) the full tight-binding model and (b) for a perturbative treatment of the Dirac Hamiltonian. In order to simulate a half-infinite graphene sheet in the numerics, we only show the region close to one of the edges of a wide nanoribbon (width $W = 154a/\sqrt{3}$). The edge state ($k = -2.4/a$) is shown for the case of next-nearest neighbor hopping with $t = 0.1t$ (black), a position-dependent staggered potential with $M_0 = -0.1t$ and $\bar{\gamma} = 0.3/a$ (red) and a position-dependent potential with $V_0 = 0.4$ (blue). The wave function on sublattice B (A) is shown as a solid (dashed) line.

Thus, the wave functions on sublattice A and B in the vicinity of the edge are indeed “linked”, as observed in the numerical results of the previous section.

In Fig. 6.13 we show a comparison of the probability density of the edge state wave function at a zigzag edge from a numerical calculation within the tight-binding model (Fig. 6.13(a)) and from Eqs (6.78), (6.79), and (6.80) obtained perturbatively within the effective Dirac Hamiltonian. Apart from the fact that the Dirac Hamiltonian in general slightly underestimates the strength of the edge state decay, we find excellent agreement between the results from the full tight-binding model and the perturbation theory of this section.

Hence, next-nearest-neighbor hopping and arbitrary position-dependent potentials and staggered potentials lead to a bending of the edge state band and an increased occupation probability on the sublattice that is not occupied without these corrections to the Hamiltonian. Although these calculations have only been performed for a single edge, we can extrapolate the results to the case of a nanoribbon of finite width W . For simplicity, we will first concentrate on the case of next-nearest neighbor hopping.

If $-\tau q W \gg 1$, i.e. when the edge state is strongly localized towards the edges of the nanoribbon, the wave function of the edge state including the effects of a finite t' can be written to a good approximation as

$$\phi(y) \cong \begin{pmatrix} \sinh(zy) + \frac{t'}{t} \sinh(z(W-y)) \\ \sinh(z(W-y)) + \frac{t'}{t} \sinh(zy) \end{pmatrix}. \quad (6.81)$$

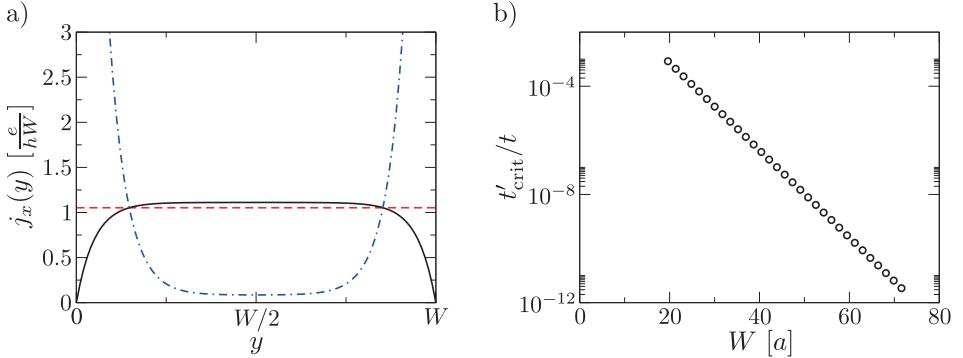


Figure 6.14 – (a) Current density $j_x(y)$ from Eq. (6.82) for $zW = 10$ and different values of t' : $t' \ll t'_\text{crit}$ (black solid line), $t' = t'_\text{crit}$ (red dashed line), and $t' \gg t'_\text{crit}$ (blue dash-dotted line). (b) Numerical calculation of t'_crit/t as a function of nanoribbon width W for $k = -2.4/a$.

Here we extrapolated Eq. (6.78) in order to formulate Eq. (6.35) for $t' > 0$: Close to the edges, the wave functions of sublattice A and B are proportional to each other, with a proportionality factor t'/t . From this wave function we can calculate the current density profile as

$$\begin{aligned} j_x(y) &\cong \frac{1}{2} \left(\cosh(zW) - \cosh(2z(W/2 - y)) \right) \\ &\quad + \frac{t'}{t} \left((\sinh(zy))^2 + (\sinh(zy))^2 \right) + \mathcal{O}((t'/t)^2) \\ &= j_{x,\text{n.n.}}(y) + \frac{t'}{t} (|\phi_A(y)|^2 + |\phi_B(y)|^2) + \mathcal{O}((t'/t)^2) \end{aligned} \quad (6.82)$$

where $j_{x,\text{n.n.}}(y)$ is the current density in nearest-neighbor approximation as given in Eq. (6.56). Thus, in addition to the current density in nearest-neighbor approximation with a maximum in the middle of the ribbon, the current density for $t' \neq 0$ contains an additional term that is *proportional to the edge state probability density*, located at the edges of the ribbon. In order to quantify which term will dominate the current density, we define the critical next-nearest neighbor hopping t'_crit such that

$$j_{x,\text{n.n.}}(W/2) = \frac{t'_\text{crit}}{t} (|\phi_A(0)|^2 + |\phi_B(0)|^2) = \frac{t'_\text{crit}}{t} (|\phi_A(W)|^2 + |\phi_B(W)|^2) , \quad (6.83)$$

i.e. the current flow at the edges equals the current flow in the middle. Then, for $t' \ll t'_\text{crit}$, current flow will be mainly through the middle of the nanoribbon, whereas for $t' \gg t'_\text{crit}$ it will be dominated by the contributions from the edge. Fig. 6.14(a) shows these different behaviors.

We can solve Eq. (6.83) for t'_{crit}/t (neglecting terms of order $(t'_{\text{crit}}/t)^2$ and higher) and obtain

$$\frac{t'_{\text{crit}}}{t} = \frac{1}{2} \frac{\cosh(zW) - 1}{\cosh(zW)^2 - 1}. \quad (6.84)$$

For $zW \gg 1$ this expression simplifies to

$$\frac{t'_{\text{crit}}}{t} \approx \frac{1}{2} e^{-zW} = \frac{1}{2} e^{\tau q W}. \quad (6.85)$$

Thus, it is enough to have a next-nearest neighbor hopping that is *exponentially small* with the width of the nanoribbon in order to change the current density profile *fundamentally*. In Fig. 6.14(b) we show the value for t'_{crit} within the tight-binding model, obtained numerically by comparing current densities at the edge and in the middle of the nanoribbon for different values of t' . Indeed, we find that t'_{crit} decreases exponentially with the width of the nanoribbon, as predicted from the effective Dirac Hamiltonian.

Note that the exact value of t' for graphene is not known: In the literature one can find values $t' \approx 0.02t' - 0.2t'$ [35, 183, 184, 190]. In this respect, it is reassuring that the critical value of t' decreases exponentially with the ribbon width: For all but the smallest ribbons, all reported values of t' will lead to the same transport properties for the edge state.

In the case of next-nearest neighbor hopping the extrapolation to the nanoribbon was particularly easy, as there the wave functions on sublattice A and B were identical up to a prefactor. This is more complicated for the other examples of this section, where there are other functions multiplying the exponential decay (Eqs. (6.79) and (6.80)). Still, eventually the exponential decay dominates as seen in Fig. 6.13, and we expect a similar behavior as for next-nearest neighbor hopping. In particular, we predict that an exponentially small perturbation will again be enough to change the edge state transport properties fundamentally.

6.4.4 Scattering from edge defects

In the previous sections we have shown that the current density profile of zigzag nanoribbons can change drastically upon certain additional terms in the nearest-neighbor tight-binding model. However, the scattering properties of the edge state are influenced not only by the current density¹²: For example, the availability of states that the edge state can scatter to is of importance. Thus, here we will demonstrate that not only the current density changes fundamentally, but also the scattering properties.

For this we consider a zigzag nanoribbon, where a single edge atom is missing at one of the edges, and calculate numerically the transmission probabilities. Depending

¹²In fact, it is not obvious how the current density enters when calculating a scattering probability. For example, in a perturbative treatment for very weak disorder, only the probability density enters (Fermi's golden rule). In contrast, here we consider strong scattering (missing atoms in the graphene lattice).

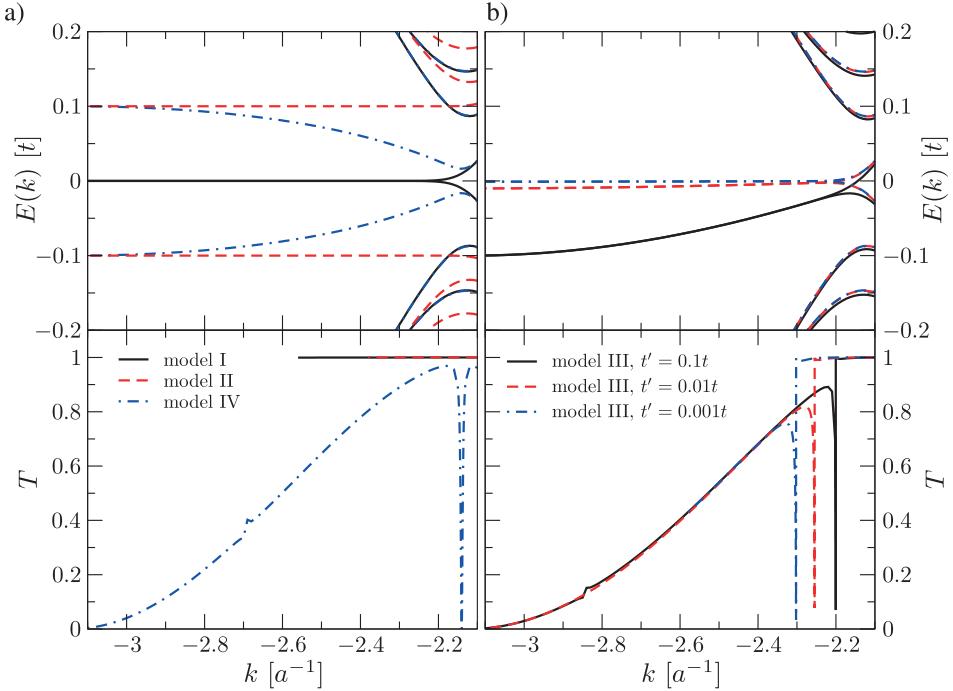


Figure 6.15 – Band structure $E(k)$ (upper panels) and transmission probability T (lower panels) as a function of the edge state wave vector k in a zigzag nanoribbon (width $W = 79a/\sqrt{3}$) where a single edge atom was removed at one of the edges. (a) shows data for nearest-neighbor hopping only (black solid line), including a constant staggered potential (red dashed line), and a position-dependent staggered potential with $M_0 = -0.1t$, $\bar{\gamma} = 1/a$ (blue dash-dotted line). (b) presents data including next-nearest-neighbor hopping for $t' = 0.1t$ (black solid line), $t' = 0.01t$ (red dashed line), and $t' = 0.001t$ (blue dash-dotted line).

on the situation, there may be more than one channel, and, in addition to that, a state localized to a single edge may be a linear combination of channels as in the case of next-nearest neighbors. Because of that we calculate the transmission eigenvalues T_n , obtained by diagonalizing tt^\dagger , where t is the transmission matrix containing all transmission amplitudes t_{mn} . The smallest transmission eigenvalue is related to the edge state, and all other transmission eigenvalues are found to be one. Thus, we simply refer to the smallest transmission eigenvalue as the “transmission probability of the edge state”.

In Fig. 6.15 we show the transmission probability as a function of the edge state wave vector k . As discussed in Section 6.3, k determines the decay constant of the edge state, and thus how strongly it is localized to the edges. For the case of nearest-

neighbors only or nearest-neighbors including a constant mass term (Fig. 6.15(a)) the transmission probability is very close to one and does not show a noticeable dependence on k , i.e. no dependence on the localization to the edges. Note that the data does not extend down to $k = -\pi/a$ in these cases: A numerical calculation cannot access a band edge, where the density of states diverges. In particular, it is not possible to reach $E = 0$ for the edge state in nearest-neighbor approximation. Even more, in nearest-neighbor approximation the energy of the edge state is exponentially small, and hence there will be a whole range of k -values around $k = \pm\frac{\pi}{a}$ that cannot be accessed numerically, because of the finite numerical precision. Thus, any perturbation that leads to an edge state dispersion is beneficial for numerical simulations in this respect, too.

Fig. 6.15(a) also shows the transmission probability for a position-dependent staggered potential. In this case, we see a pronounced dependence on k : As the edge state localizes stronger to the edge, the probability of scattering increases. In addition to this monotonic behavior, we observe some other features related to the band structure. In particular, the transmission probability is strongly suppressed (the edge state scatters strongly from the defect) at the extremum of the edge state band, close to the K -point ($k \approx -2.15/a$). The extremum of the edge state band separates left- and right-going states (corresponding to a positive and negative slope of the energy band), as does any band minimum or maximum. Close to this point, the wave functions of left- and right-going states have a good overlap, leading to a large probability of back-scattering from the defect. In addition, the back-scattering occurs within a single K -valley in this case. Aside from this pronounced effect, we also observe an additional small “wiggle” in the transmission probability ($k \approx -2.7/a$). This occurs whenever a new channel opens, as seen by comparing the respective energies with the plot of the band structure in Fig. 6.15(a).

For finite next-nearest neighbor hopping we observe an analogous behavior, shown in Fig. 6.15(b). In this case, the dependence of the transmission probability on k is almost independent of the exact value of t' , except for a shift in the position of the edge state band extremum. Note that t' covers two orders of magnitude in the examples shown in Fig. 6.15(b). This may be explained by the fact that the shape of the current density is largely independent of t' , as soon as $t' \gg t'_{\text{crit}}$. t' only determines the energy window where the edge state exists, as seen from the band structure plots in Fig. 6.15(b). This is another example demonstrating that the nature of the edge state is better characterized by the wave vector k , rather than the energy. Furthermore, although the calculated transmission probabilities for next-nearest neighbor hopping and the position-dependent staggered potential differ quantitatively, they show the same qualitative behavior.

Hence, next-nearest neighbor hopping or a position-dependent (staggered) potential not only leads to a current distribution located at the edges, but also results in substantial scattering from edge defects. The scattering is stronger, the more the edge state is localized to the edge.

6.5 Summary

In this chapter we have presented an exhaustive study on the properties of the graphene edge state. Within the commonly used nearest-neighbor tight-binding approximation for graphene, the edge state in a zigzag nanoribbon always occupies both edges, and hence both edges are maximally correlated. Moreover, the current carried by the edge state is mainly located in the center of the ribbon, opposite to the charge density localized at both sides of the ribbon. Thus, the edge state is affected only little by defects at the nanoribbon edges. With respect to transport, the edge state in this approximation rather behaves like a bulk state. These findings are in agreement with previous studies [177–180]. Introducing a constant staggered potential localizes the state to a single edge only, but does not change the transport properties. In this case we find the paradoxical situation that the charge density is asymmetric (localized at a single edge only), but the current density is symmetric with respect to the ribbon axis (flowing through the middle of the ribbon).

However, when next-nearest neighbor hopping, a position-dependent potential or a position-dependent staggered potential is included, the transport properties of the edge state change fundamentally: The current density is now also localized to the edges, in accordance with the charge density. As a consequence, the edge state scatters substantially from edge defects. This scattering is stronger, the more the state is localized to the edge. Furthermore, the edge state at either side of the nanoribbon is not affected by the opposite edge, and hence both edges are uncorrelated. Moreover, these fundamental changes prevail even if the perturbation induced by the above-mentioned terms is exponentially small with the nanoribbon width. Thus, the properties of the edge state that were complicated in a simple model of graphene, namely nearest-neighbor tight-binding, become simple and intuitive in a more complex model.

The fact that current and charge density are drastically different for zigzag nanoribbons in nearest-neighbor tight-binding approximation leads to many counter-intuitive effects and is unique to graphene. Yet, already an exponentially small perturbation leads to a fundamentally different behavior of the edge state. Hence, the nearest-neighbor tight-binding approximation is a rather singular model with respect to the edge state, and does not apply to any realistic system.

Chapter 7

Edge state based spintronics in graphene

7.1 Introduction

Amongst its many unusual electronic properties, graphene is also a good candidate for spintronics applications: The spin relaxation time in graphene is expected to be very long, since spin-orbit effects are expected to be weak [39–41, 191], and because the predominant isotope in natural carbon is C¹² which has zero nuclear spin, resulting in a vanishing hyperfine interaction [38]. In fact, spin injection from ferromagnetic metal contacts into graphene and coherent spin transport in graphene have been already demonstrated experimentally [42–46].

For ohmic contacts to the ferromagnetic electrodes, the conductance mismatch between the three-dimensional ferromagnetic metal contacts and the two-dimensional graphene sheet prevents efficient spin injection [192]. In order to overcome this obstacle in the experiment, a tunnel barrier is introduced between the ferromagnetic contact and the graphene sheet as suggested in Refs. [193, 194]. Thus, using ferromagnetic metals as a source of spin-polarized electrons for graphene involves a complicated layered structure with very different materials: Ferromagnetic metals as electrodes, some insulating material (Al₂O₃ in Refs. [43, 45, 46]), and finally graphene including the underlying substrate. In order to avoid the conductance mismatch problem, it would be desirable to arrive at a spin-injection design based on graphene only.

The graphene edge state at a zigzag edge is a promising candidate for magnetism, as it was predicted to exhibit a locally ferrimagnetic state [170]: Spins within the same sublattice are aligned parallel, whereas spins on different sublattices are aligned antiparallel. As a result, a zigzag nanoribbon shows a magnetic structure, where the spins at one of the edges are predominantly aligned in one direction, and the spins

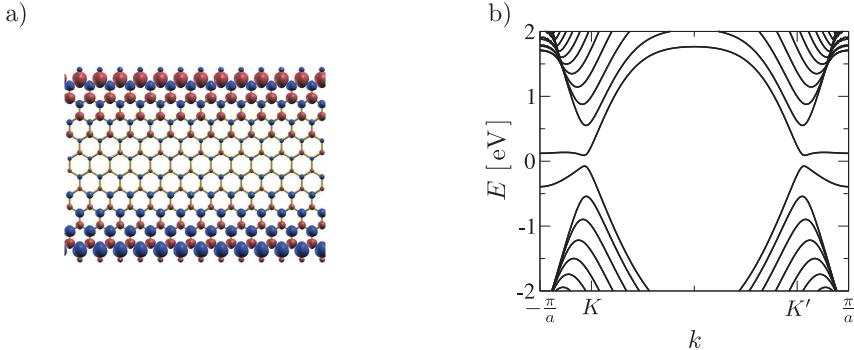


Figure 7.1 – (a) Ground state spin density of a zigzag graphene nanoribbon as obtained from DFT [195]. Spin up (down) is shown in blue (red). (b) Band structure of a zigzag graphene nanoribbon as obtained from DFT [196].

on the opposite edge in the antiparallel direction, as shown in Fig. 7.1(a). These predictions have been confirmed by density functional theory (DFT) calculations [164, 175, 176, 197], but although the existence of edge states has been verified through scanning tunneling spectroscopy, the existence of edge magnetism has not been shown experimentally yet. The antiferromagnetic ordering between the different sublattices opens a gap in the band structure of the zigzag nanoribbon, as shown in Fig. 7.1(b). Hence, edge magnetism is one possible explanation for the experimentally observed band gap of graphene nanoribbons (GNRs) [158–160].

The magnetic properties of the edge state have been studied in great detail in several theoretical works. For example, the antiferromagnetic ordering between the different sublattices can be understood from a theorem about the Hubbard model on a bipartite lattice [198]. In addition, the ferromagnetic ordering at a single edge can be understood as a consequence of the Stoner criterion [199], as the edge states give rise to a very high density of states (see the discussion in the previous chapter). Still in general, a high density of states at the Fermi energy does not necessarily lead to magnetism, but may also result in some other instability, such as a geometrical distortion. However, DFT calculations favor magnetism over other instabilities [197]. Furthermore, from a general theorem that excludes ferromagnetic or antiferromagnetic ordering in one-dimensional systems [200] it may seem that a zigzag graphene edge cannot exhibit true long-range magnetic ordering. However, long-range order has been shown to prevail at experimentally relevant, finite temperatures due to intrinsic or substrate-induced spin-orbit effects [201]. In addition to that, edge magnetism has been shown to exist not only in clean graphene nanoribbons, but also at disordered edges, provided that zigzag pieces with a length of at least 3–4 unit cells exist [195, 202–205]. Moreover, since edge states are expected to exist also for edge orientations other than zigzag [171],

edge magnetism also prevails in nanoribbons that are tilted with respect to the zigzag direction [204]. Finally, in order to achieve transport, the zigzag GNR must be doped to bring the Fermi energy into a regime with open conduction channels. This can be achieved in practice by a gate voltage or chemical doping. As the Fermi energy moves away from the high density of states created by the edge states, the magnetization decreases—only the edge states are spin-polarized, not the bulk states—until finally the magnetic ordering is destroyed. Within DFT calculations, the critical value of this doping is found to be ≈ 0.5 electrons (≈ 0.4 holes) per zigzag edge atom [195]. Hence, the magnetic structure survives a finite amount of doping necessary for transport.

In summary, many different theoretical predictions strongly support the existence of edge magnetism. In the remainder of this chapter we will therefore discuss the implications of this edge magnetization for graphene spintronics. Due to the nature of the magnetic ordering—antiferromagnetic between the two sublattices—a zigzag GNR does not exhibit global, but only local spin-polarization. Hence, the edge magnetism in clean GNRs cannot be used directly to generate spin-polarized currents. The objective of this chapter is to show new concepts how to generate spin-polarized and pure spin currents using the edge magnetism in zigzag graphene nanoribbons.

To this end, the chapter is organized as follows: First, we introduce a mean-field model to include the edge magnetism in the tight-binding model for graphene. Then, we show that zigzag graphene nanoribbons with rough edges exhibit universal spin conductance fluctuations and are thus a natural source of spin-polarized electrons. Based on these results we suggest an experiment to prove the existence of edge magnetism, using an all-electrical measurement. Finally, we show that the edge state can lead to a spin Hall effect in ballistic structures. In particular, this spin Hall effect can be used to generate spin-polarized currents in a three-terminal device, and a pure spin current in a four-terminal geometry. Since spin-polarized and pure spin currents are a prerequisite for spintronics applications, these results may serve as a foundation for an all-graphene spintronics.

7.2 Mean-field model for edge-state magnetism

We describe the edge magnetism in mean-field approximation within the tight-binding model for graphene. In this approximation, the Hamiltonian contains an additional term [170]

$$H_{\text{mag}} = \sum_{i,\alpha} \frac{1}{2} \mathbf{M}_{i\alpha} \cdot \mathbf{s} |i\alpha\rangle \langle i\alpha| , \quad (7.1)$$

where again (i, α) denotes the lattice site ($\alpha = A, B$), $\mathbf{M}_{i\alpha}$ the local magnetization and \mathbf{s} the vector of Pauli matrices in (real, not pseudo-) spin space. The local magnetization can be determined self-consistently within the mean-field approximation to the Hubbard model [170], or obtained from DFT calculations. In this work however, we choose to employ a phenomenological model for $\mathbf{M}_{i\alpha}$, capturing the essential physics.

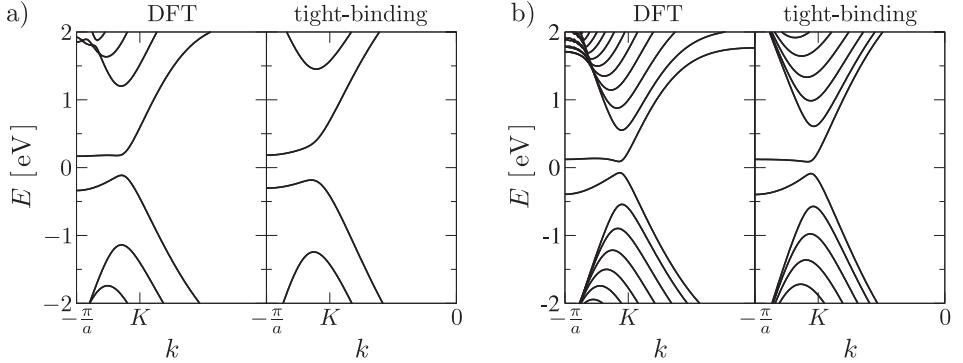


Figure 7.2 – Comparison between the band structure of a zigzag graphene nanoribbon, as obtained from DFT and the mean-field model within the tight-binding approximation. (a) shows the band structures of a GNR with width $W = 13a/\sqrt{3}$ and for the constant magnetization approximation with $M = 0.18t$. (b) Band structures for a ribbon with width $W = 31a/\sqrt{3}$ and the position-dependent magnetization from Eq. (7.4) with $M_0 = 0.192t$ and $t' = 0.067t$.

As seen in the previous section, spins are aligned in parallel within a single sublattice, but antiparallel between different sublattices. Hence, we can write the magnetic term in the Hamiltonian as

$$\langle i\alpha | H_{\text{mag}} | i\alpha \rangle = \begin{cases} +\frac{M_i}{2} s^z & \text{for } \alpha = A, \\ -\frac{M_i}{2} s^z & \text{for } \alpha = B. \end{cases} \quad (7.2)$$

Since the local magnetization forms a collinear texture, we can choose the magnetization to be in z -direction without loss of generality. Thus, the system can be decomposed into two independent subsystems for spin up and spin down. Within these subspaces, the magnetization term reduces to a position-dependent staggered potential, with opposite sign for spin up and spin down. Within the effective Dirac Hamiltonian, the magnetic term takes the form

$$H_{\text{mag}} = \frac{1}{2} M(\mathbf{x}) s^z \sigma^z, \quad (7.3)$$

where again σ^z acts in pseudospin (sublattice) space.

Finally, we have to specify the position dependence of the magnetization M_i . In the simplest approximation, we neglect the position dependence and set $M_i = M$, where M is a constant magnetization. Using the tight-binding model in next-nearest neighbor approximation with $t = 2.7 \text{ eV}$ and $t' = 0.1t$ we fit the magnetization M to the results of the DFT calculations for a narrow nanoribbon with width $W = 13a/\sqrt{3}$. We find a very good agreement between the band structures of the tight-binding model and the DFT calculations for $M = 0.18t$, as shown in Fig. 7.2(a).

constant M	$t = 2.7 \text{ eV}$	$t' = 0.1t$	$M = 0.18t$
position-dependent M_i (Eq.(7.4))	$t = 2.7 \text{ eV}$	$t' = 0.067t$	$M_0 = 0.192t$

Table 7.1 – Mean-field model parameters fitted to *ab-initio* DFT calculations.

A more realistic model, that also seeks to describe all details of the band structure in wider ribbons, must take into account the position dependence of the magnetization. Since the magnetization originates from spin-polarized edge states, it is largest at the edges and decays towards the middle of the graphene nanoribbon. We can estimate the magnetization close to an edge by summing over all edge state solutions, Eq. (6.22), such that

$$M_j = M_0 \times \frac{3a}{2\pi} \int_{\frac{2\pi}{3a}}^{\frac{4\pi}{3a}} dk |\phi_{\text{edge}}(k, j)|^2 = M_0 \times \frac{3a}{2\pi} \int_{\frac{2\pi}{3a}}^{\frac{4\pi}{3a}} dk e^{2\gamma(k)j}, \quad (7.4)$$

where j is an index numbering the atoms relative to the edge, and $\gamma(k)$ is the decay constant of the edge state, that depends on the lateral Bloch vector k .¹ The integral has been normalized such that the magnetization has the value M_0 at the edge atom.

We now fit both t' and M_0 to the results of DFT calculations for a nanoribbon with width $W = 31a/\sqrt{3}$, which is the widest nanoribbon for which we have access to DFT data. For $t' = 0.067t$ and $M_0 = 0.192t$ we find very good agreement between the band structures from DFT and the tight-binding model, as shown in Fig. 7.2(b). At $k = \pm\pi/a$ the band structures exhibit a gap of size M_0 , since the edge state is then localized on the outermost atoms only. Moving closer to the K and K' -point, the band gap decreases. In this situation the edge states decay further into the nanoribbon and thus feel an average magnetization smaller than M_0 .

The parameters for the two models are summarized in Table 7.1. As we show below, the characteristics of the spin-polarized edge state transport does not depend on the details of the position dependence of the magnetization. The position dependence only modifies the energy window, where transport is dominated by the spin-polarized edge states. For this reason, we will employ the constant magnetization approximation also for wide ribbons, unless explicitly stated otherwise.

It is important to note that in both models the transport properties of the edge state are qualitatively the same: In both cases, current density is highest at the edge, and the edge state transport is strongly influenced by edge impurities, as discussed in Chapter 6. In contrast, if we had chosen a constant magnetization but $t' = 0$, the transport properties characteristics would be radically different. Such a singular model of graphene transport would also lead to radically different spin transport properties, as we demonstrate below.

¹ Note that we used the $2\pi/a$ -periodicity of the Bloch vector k in order to write the integral in a compact form.

7.3 Spin currents in rough graphene nanoribbons

7.3.1 Basic ideas

In the following, we work in the linear response regime and assume that the phase coherence length is much larger than the system size. The *spin conductance* of a zigzag GNR is defined as

$$G_s = \frac{e}{4\pi} (T_{\uparrow} - T_{\downarrow}), \quad (7.5)$$

where $T_{\uparrow(\downarrow)}$ is the transmission probability for spin up (down). Note that since the magnetization is collinear, the system decomposes into two independent subsystems, and no spin flip can occur. In contrast to the spin conductance, the total (charge) conductance of the system is given as (Eq. (2.120))

$$G_{\text{tot}} = \frac{e^2}{h} (T_{\uparrow} + T_{\downarrow}). \quad (7.6)$$

Note that the units of spin conductance are such that the spin current has units angular momentum per time, whereas the charge current has units charge per time.

Since the edge magnetization has the form of a staggered potential, the edge state is localized to a single edge only, as discussed in Section 6.3.3. Since spin up and down feel a staggered potential of opposite sign, Eq. (7.2), the different spins are localized to opposite edges, as shown in Fig. 7.3(a).

Due to the spatial localization of the edge state towards the zigzag edge, the edge state is scattered strongly by defects at the same edge, whereas it remains unaffected by defects on the opposite edge². Hence, distinguishing a left (l) and right (r) edge of the nanoribbon, we can thus approximate $T_{\uparrow(\downarrow)}$ by $T_{l(r)}$, where $T_{l(r)}$ is the transmission probability of the corresponding edge state assuming that the opposite edge is not disordered. The transport properties of the zigzag GNR are thus essentially those of two independent one-dimensional wires, oppositely spin polarized and connected in parallel between the reservoirs. We will demonstrate the validity of this two-wire model using numerical calculations.

In a clean zigzag graphene nanoribbon, the symmetry between both edges implies $T_l = T_r$ such that the spin conductance of a perfect zigzag GNR vanishes. This is demonstrated in Fig. 7.3(a): both spins are transmitted equally into the doped bulk graphene³. In order to obtain a finite spin conductance, we need to break the symmetry between both edges. This can be achieved by a large transverse electric field, as shown in Ref. [176]. In this case, one of the edge states is shifted out of the Fermi window,

² As discussed in Chapter 6, this is the case for any realistic model of graphene. Here, “realistic” means going beyond the nearest-neighbor tight-binding approximation that behaves rather singular with respect to edge state transport

³ Although the nanoribbon has zero spin conductance in this situation, there may be a spin Hall effect—as seen in Fig. 7.3(a), different spins are deflected in opposite directions. We will discuss this effect in detail in Section 7.4.

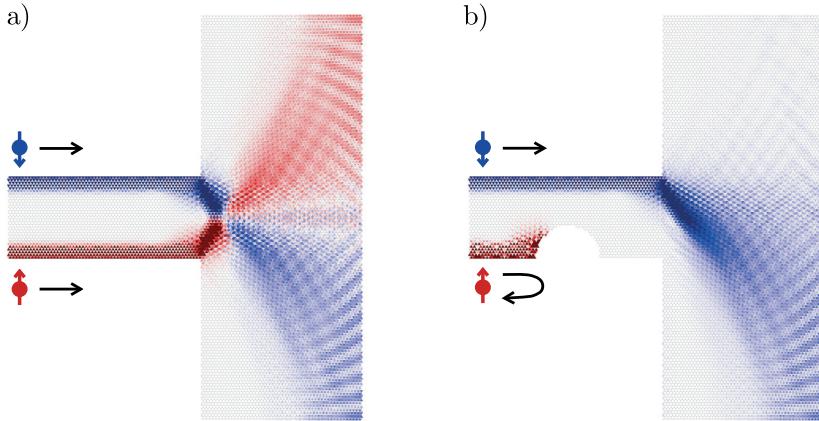


Figure 7.3 – Spin injection from (a) a perfect zigzag GNR and (b) a zigzag GNR with a large obstacle at one edge. Spin up (down) non-equilibrium densities (Eq. (A.22)) are shown in red (blue). In order to simulate spin injection into bulk graphene, where many transport channels are available, the band edge of the wider graphene ribbon has been shifted by a potential $V_{\text{gate}} = -0.6t$, mimicking the effect of a gate voltage. Note that the interference fringes are due to back reflection from the edge of the wide ribbon.

rendering the zigzag GNR completely spin-polarized. Here we will focus on alternative symmetry breaking method, that may be of more importance for the experiment: disorder at the graphene edges. If one of the edges is disordered more strongly, say the right edge, we can expect $T_r < T_l$, leading to a finite spin conductance. This is demonstrated in Fig. 7.3(b) where a large obstacle at one edge blocks the corresponding spin channel completely, leading to perfect spin injection. Hence, by controlling the roughness of the edges, zigzag GNRs can be designed to be a source of spin-polarized electrons.

7.3.2 Spin conductance of rough graphene nanoribbons

From an experimental perspective, unless nanoribbons are specifically fabricated with edges of different roughness, both edges will be similarly disordered. In this case, the average transmission probability of both edges will be equal $\langle T_l \rangle = \langle T_r \rangle$, resulting in a vanishing average spin conductance

$$\langle G_s \rangle = 0. \quad (7.7)$$

Here, $\langle \dots \rangle$ denotes an average over edge disorder configurations. However, since the disorder on both edges of the GNR is usually uncorrelated, a *single* graphene nanorib-

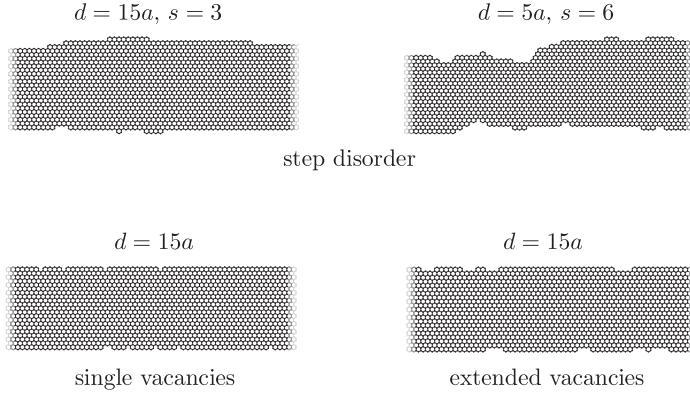


Figure 7.4 – Overview of the different disorder models used in this work. Step disorder: edge disorder created by a random walk, where the width of the nanoribbon is changed by one hexagon at every step. Steps are made with probability a/d and the maximum deviation of the width is $\leq s$ hexagons. Single vacancies: edge atoms are removed randomly with probability a/d . Extended vacancies: similar to single vacancies, but also neighboring edge atoms are removed. In all cases the disorder on the two edges is uncorrelated.

bon will typically exhibit $T_l \neq T_r$ for a given energy, leading to mesoscopic spin conductance fluctuations. Within the two-wire model we find the variance $\text{Var } G_s$ of the spin conductance as

$$\text{Var } G_s = \left(\frac{e}{4\pi}\right)^2 (\text{Var } T_l + \text{Var } T_r) = \left(\frac{h}{2e}\right)^2 \text{Var } G_{\text{tot}}. \quad (7.8)$$

Since the transmission probabilities of both edges are uncorrelated, the variance of the spin conductance is simply given by the sum of the variances of the transmission probabilities for both edges. Hence, the variance of the *spin* conductance is equal to the variance of the *charge* conductance, up to a factor that is only due to the different units of spin and charge conductance.

In order to demonstrate the validity of the predictions of the two-wire model, we now perform numerical simulations of zigzag GNRs of different length L and width W . We employ several models of edge roughness, as summarized in Fig. 7.4, in order to show the independence of the effects on the type of edge disorder. In all disorder models, d denotes the average distance between scatterers. We first focus on the case of dilute disorder, $d \gg a$, and come back to the problem of strong disorder, $d \approx a$, in the next section.

We start by considering the ensemble averaged properties of a zigzag GNR, i.e. the average conductance and its fluctuations. Fig. 7.5(a) shows the typical behavior of the charge and spin conductances. In order to compare the average conductance and

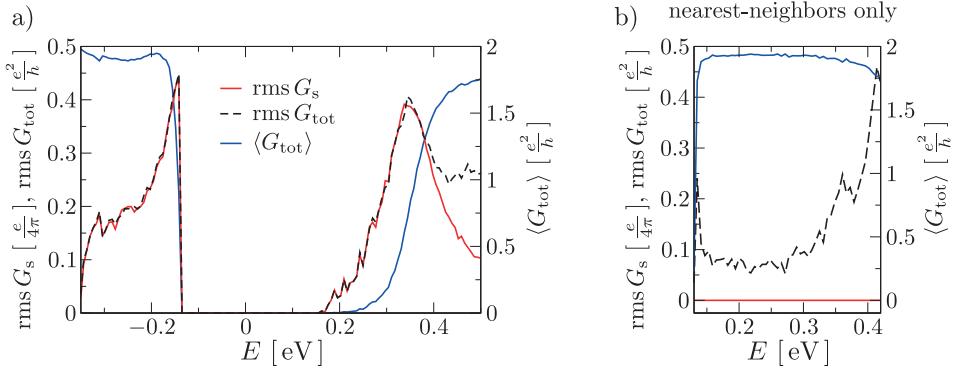


Figure 7.5 – (a) Average total conductance, $\langle G_{\text{tot}} \rangle$ (blue solid line), rms of the total conductance, $\text{rms } G_{\text{tot}}$ (black dashed line), and rms of the spin conductance, $\text{rms } G_s$ (red solid line), as a function of E ($E = 0$ is chosen to correspond to zero gate voltage). The data were averaged over 1000 configurations of single vacancies with $d = 40a$, for a GNR with $W = 34a/\sqrt{3}$ and $L = 800a$. (b) shows the same quantities for a model with constant magnetization and nearest-neighbor hopping only ($t' = 0$).

the conductance fluctuations quantitatively, we characterize the fluctuations by the root-mean-square conductance, $\text{rms } G = \sqrt{\text{Var } G}$.

Note that over the whole energy region where the edge states are present we observe $\frac{\hbar}{2e} \text{rms } G_{\text{tot}} \approx \text{rms } G_s$, confirming the validity of the two-wire model. As the Fermi level is raised by gating or doping, the edge states extend further into the ribbon and start to feel both edges. Then, the assumption of uncorrelated channels breaks down, and $\frac{\hbar}{2e} \text{rms } G_{\text{tot}} > \text{rms } G_s$.

For an n-type GNR, when the Fermi level is near the band edge, the states are localized and both the average conductance and the fluctuations are suppressed exponentially. Raising the energy E , we observe in Fig. 7.5(a) a crossover to the ballistic regime, where the conductance rises up to the quantum limit of conductance $2e^2/h$. Correspondingly, we see a maximum in the conductance fluctuations before they vanish again in the ballistic regime.

In the p-type regime, the situation is more complicated. Due to the finite dispersion of the edge state band, there are two open channels at the Fermi level (see Fig. 7.2(a)). One of these states is an edge state and strongly localized to the edge, whereas the other one extends further into the ribbon. As discussed in Section 6.4.4, scattering is strongest at the band edge, and correspondingly we observe a maximum in the spin conductance fluctuations close to the band edge. Although the n- and p-type regimes are seemingly quite different, they can be both described in a unified way, as we show in the next section.

Finally, we also consider the charge and spin conductance fluctuations for graphene

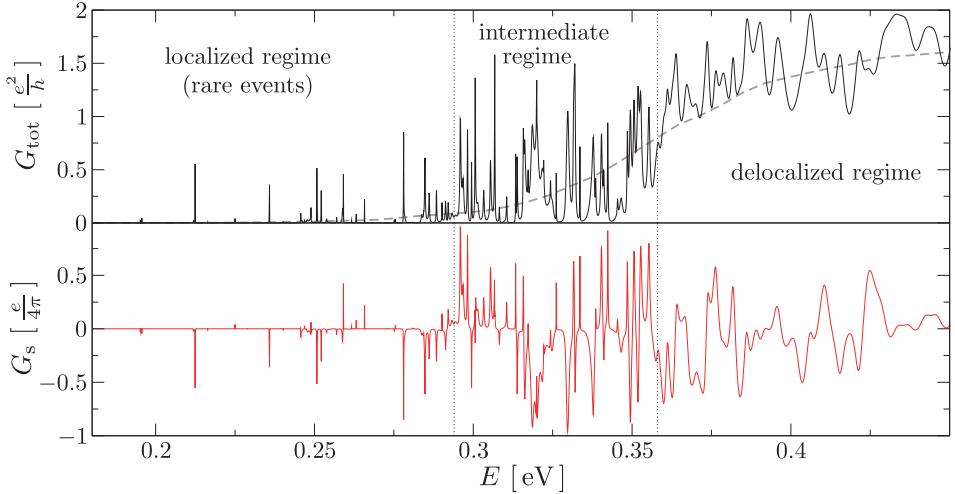


Figure 7.6 – (a) Charge conductance G_{tot} (black solid line) and (b) spin conductance G_s (red solid line) as a function of energy E for a single graphene nanoribbon with $W = 34a/\sqrt{3}$ and $L = 800a$. As in Fig. 7.5, the edge disorder model is single vacancies with $d = 40a$, but instead of an average over several disorder configurations, we only show the transport properties of a single configuration. For comparison, the grey dashed curve in (a) shows $\langle G_{\text{tot}} \rangle$ from Fig. 7.5(a).

in nearest-neighbor tight-binding approximation with a constant magnetization. In Chapter 6 we have shown that this is a rather singular model with respect to edge state charge transport. In particular, the edge state was influenced equally by scatterers on both edges. Fig. 7.5(b) shows that this also leads to a singular behavior with respect to spin transport properties: Whereas we obtain finite charge conductance fluctuations, the spin conductance fluctuations vanish. Since the edge state is influenced by both edges equally, $T_\uparrow = T_\downarrow$, and the spin conductance is always zero. Hence, also for spin transport properties it is essential to go beyond the nearest-neighbor tight-binding approximation.

After discussing ensemble-averaged properties, we now turn to the transport properties of a single, disordered nanoribbon. Experimentally, the average conductance would be observed when measuring an array of different nanoribbons at the same time. In this case, spin conductance would be quenched. However, if a single GNR is measured, the results will be dominated by the conductance fluctuations.

In Fig. 7.6 we show both the charge and the spin conductance of a single GNR as a function of energy E in the n-type regime. In order to compare the average charge conductance $\langle G_{\text{tot}} \rangle$ with the charge conductance of a single ribbon, we show both quantities in the upper panel of Fig. 7.6. We can clearly identify different regimes:

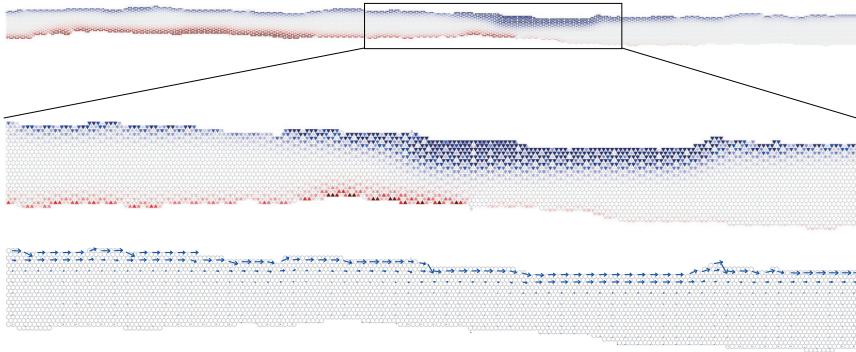


Figure 7.7 – Non-equilibrium spin (upper and middle panel) and current density (lower panel) of a zigzag GNR with $W = 34a/\sqrt{3}$ and $L = 600a$. The figure shows one configuration of a GNR with step disorder ($d = 10a$ and $s = 6$). In the upper two panels, Spin up (down) density is shown in blue (red). The lower two panels show a magnified part of the nanoribbon, as indicated in the upper panel.

For low energies, close to the band edge, the system is localized and the conductance is dominated by rare events that appear as sharp peaks in the transmission function with a large transmission probability. These are due to the long tail of the distribution function for the transmission probabilities, and the average conductance is not a representative quantity [206]. Remarkably, every one of these rare events leads to a large spin conductance. Since these large transmission probabilities are “rare”, only one of the spin-polarized edges will transmit, whereas the other one is blocked, leading to a completely spin-polarized current. For large energies, where the edge state starts to extend further into the ribbon and scattering from the edge becomes weaker, the conductance of the single GNR follows the average conductance, albeit with sizable smooth fluctuations. These fluctuations somewhat resemble the universal conductance fluctuations (UCF) known from quasi-one-dimensional wires [207] with many channels. Strictly speaking, this is not true: The conductance arises from a single channel only at each edge, thus acting as a one-dimensional wire. A one-dimensional system however always localizes, and a metallic regime is not present [206, 207]. Again, these charge conductance fluctuations give rise to (smooth) spin conductance fluctuations. In the intermediate regime we observe a mixture of smooth fluctuations and rare events, that again carries over to the spin conductance. In fact, as seen from Fig. 7.5(a), the spin conductance fluctuations are largest in this case.

Finally, we illustrate the wave function associated with a rare event in Fig. (7.7). There, we chose to plot the non-equilibrium spin density and the charge current for an energy E where the transmission probability is strongly peaked with $T(E) \approx 1$. The edge state “meanders” along the strongly disordered edge, as seen both from the

spin as well as current densities, that exactly follow the shape of the edge. While one of the edges transmits with a high probability, the edge state on the opposite side is eventually blocked, leading to a completely spin-polarized current. In fact, it is quite remarkable that the edge state, which is strongly localized towards the edge in this case, can follow variations of the ribbon edge larger than its own extent. This is due to the fact that a few hexagons in zigzag orientation are enough to support a local edge state [169]; in order to transmit, the edge state then only has to overcome a single step to arrive at the next available state. We will comment on this topic some more at the end of the next section.

7.3.3 Universal spin conductance fluctuations

After discussing the typical behavior of the spin conductance and verifying the predictions of the two-wire model for a given disorder model, we will now show that the properties of the spin conductance are in fact independent on the model for edge disorder. In particular, we will find a universal maximum value of the spin conductance fluctuations.

Within the two-wire model, we have stated that the conductance of a zigzag GNR is given by the conductances of two independent, disordered wires. The transmission eigenvalue statistics of disordered wires is governed by the Dorokhov-Mello-Pereyra-Kumar (DMPK) equation [208, 209] (for a review, see Ref. [207]). For the one-dimensional, single-channel case, the DMPK equation reduces to

$$l \frac{\partial}{\partial L} P(\lambda, L) = \frac{\partial}{\partial \lambda} \lambda (1 + \lambda) \frac{\partial}{\partial \lambda} P(\lambda, L), \quad (7.9)$$

where l is the mean free path, L the wire length, $\lambda = (1 - T)/T$, T the transmission probability and $P(\lambda, L)$ a probability distribution function for λ , normalized such that $\int_0^\infty d\lambda P(\lambda, L) = 1$. From this distribution function, we can calculate the expectation value of any function of the transmission probability, and this value is uniquely determined by the ratio L/l of the mean free path l and the length of the wire L .⁴ This fact is also referred to as single-parameter scaling.

Note that the differential equation (7.9), that is an equation for the distribution function, can be converted to a differential equation for expectation values. This is particularly useful for

$$\langle \ln T \rangle = - \int_0^\infty d\lambda \ln(1 + \lambda) P(\lambda, L), \quad (7.10)$$

⁴ This can be seen by a change of coordinates $L \rightarrow L/l$.

as this results in a closed differential equation:

$$\begin{aligned}
l \frac{\partial}{\partial L} \langle \ln T \rangle &= - \int_0^\infty d\lambda \ln(1+\lambda) l \frac{\partial}{\partial L} P(\lambda, L) \\
&= - \int_0^\infty d\lambda \ln(1+\lambda) \frac{\partial}{\partial \lambda} \lambda (1+\lambda) \frac{\partial}{\partial \lambda} P(\lambda, L) \\
&= + \int_0^\infty d\lambda \lambda \frac{\partial}{\partial \lambda} P(\lambda, L) \\
&= - \int_0^\infty d\lambda P(\lambda, L) = -1,
\end{aligned} \tag{7.11}$$

where we used Eq. (7.9), the normalization of $P(\lambda, L)$ and integration by parts twice. Together with the boundary condition $\lim_{L \rightarrow 0} \langle \ln T \rangle = 0$ (i.e. $T \rightarrow 1$) we arrive at

$$\langle \ln T \rangle = - \frac{2L}{\xi}, \tag{7.12}$$

where $\xi = 2l$ is the *localization length* [207].

Apart from giving rise to differential equations for expectation values, the DMPK-equation (7.9) can also be solved for the distribution function $P(\lambda, L)$ directly (see Refs. [210, Eq. (20)] and [211, Eq. (4.3)]):

$$P(\lambda, L) = \frac{1}{\sqrt{2\pi}} \left(\frac{l}{L} \right)^{\frac{3}{2}} e^{-\frac{L}{4l}} \int_{\text{arccosh}(1+2\lambda)}^\infty du \frac{u e^{-\frac{u^2 l}{4L}}}{\sqrt{\cosh u - 1 - 2\lambda}}. \tag{7.13}$$

From this explicit form of the distribution value, we can calculate arbitrary expectation values—at least numerically. In particular, we obtain the variance of the spin conductance within the two-wire model as

$$\text{Var } G_s = \left(\frac{e}{4\pi} \right)^2 2 \text{Var } T = \frac{2e^2}{(4\pi)^2} \left(\int_0^\infty d\lambda \frac{P(\lambda, L)}{(1+\lambda)^2} - \left(\int_0^\infty d\lambda \frac{P(\lambda, L)}{1+\lambda} \right)^2 \right). \tag{7.14}$$

As seen from the DMPK equation (7.9), $\text{Var } G_s$ only depends on l/L , or equivalently ξ/L . Hence, in order to compare the spin conductance fluctuations for different disorder models and n- as well as p-type ribbons, we extract the energy dependence of the localization length $\xi(E)$ from the numerical simulations using Eq. (7.12). The result for different disorder models is shown in Fig. 7.8. We observe a monotonic increase of the localization length ξ with increasing (decreasing) energy in the n-type (p-type) regime, as the edge state extends further into the ribbon and is scattered less from edge defects. Note that the disorder involving single vacancies only (red dashed curve in Fig. 7.8) has the largest distance d between scatterers, but on average the shortest localization length. At first glance this is somewhat unexpected, but can be understood from the fact that all other types of disorder always feature at least a few hexagons in zigzag orientation in a row, supporting a local edge state. This is not the case for

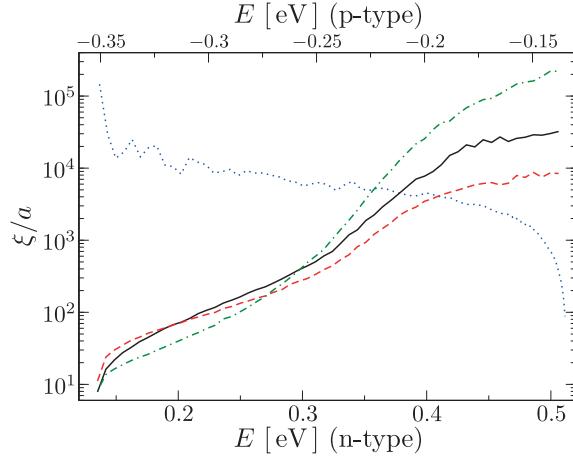


Figure 7.8 – ξ/a as a function of E for step disorder ($d = 30a$, $s = 3$; black solid line), single vacancies ($d = 40a$; n-doped: red dashed line, p-doped: blue dotted line) and extended vacancies ($d = 30a$, green dashed-dotted line).

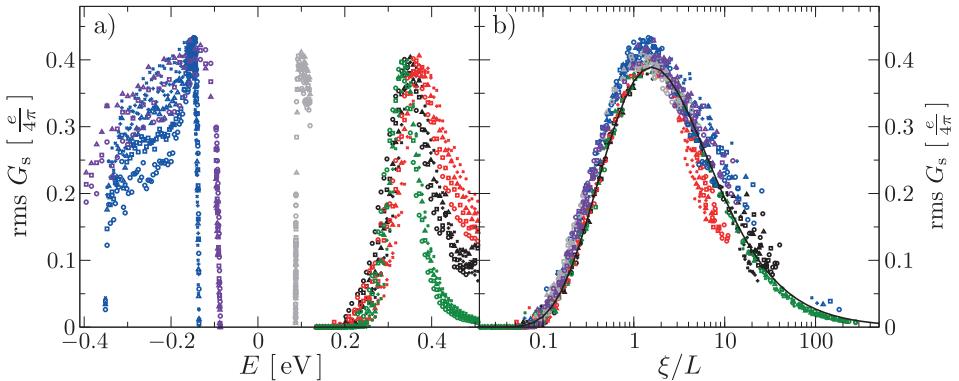


Figure 7.9 – Spin conductance fluctuations: rms G_s as a function of (a) E and (b) ξ/L for n- and p-doped graphene. We show data within the constant magnetization approximation: step disorder for n-type, $d = 20a$, $s = 3$ (black), single vacancies for n- and p-type, $d = 40a$ (red and blue, respectively) and extended vacancies for n-type, $d = 30a$ (green); we also present data for the model including a position-dependent magnetization: single vacancies for n- and p-type, $d = 40a$ (grey and violet, respectively). The solid line corresponds to the DMPK prediction, Eq. (7.14). The data is shown for graphene nanoribbons with lengths $L = 800a$ (\circlearrowleft), $1000a$ (\square), $1200a$ (\triangle), $1400a$ (+), and $1600a$ (\times), and width $W = 34/\sqrt{3}a$. The rms G_s is estimated from 1000 disorder configurations.

the single vacancies which only involve one hexagon in the graphene lattice, and thus leads to stronger scattering.

In Fig. 7.9 we show rms G_s for zigzag GNRs within the constant magnetization approximation and a more realistic model with a position-dependent magnetization (as discussed in Section 7.2) and all three disorder models (as shown in Fig. 7.4), with different values of d and a wide range of ribbon lengths L . Whereas the functional dependence of the rms conductance rms G_s on the energy E is very different for the various models as shown in Fig. 7.9(a), all the data collapse onto a single curve when plotted as a function of ξ/L (Fig. 7.9(b)). The results of the numerical simulations agree very well with the DMPK prediction in Eq. (7.14) (shown as black solid line in Fig. 7.9(b)). The agreement is especially good for n-doped GNRs in constant magnetization approximation, where there is only one conductance channel per spin. In this case, the single-channel DMPK equation (7.9) holds exactly. For the p-doped ribbon in constant magnetization approximation and the ribbons with a position-dependent magnetization, there are two conductance channels per spin (c.f. Fig. 7.2), and we observe a slight increase in the rms conductance with respect to the single-channel case. In this situation, the predictions of the single-channel DMPK equation only hold approximately, and the increase in rms G_s is presumably due to a beginning crossover to a multi-channel quasi-1D wire [207]. For all models, the maximum value of the spin conductance fluctuations fits well to rms $G_s \approx 0.4e/4\pi$. These considerations show that the spin conductance fluctuations are indeed independent from the model employed to describe the edge magnetization and the type of disorder, and are universal in this respect.

In Fig. 7.10 we concentrate on n-doped graphene for step disorder (upper panels of Fig. 7.4), that may be closest to a real, experimentally obtained nanoribbon. Fig. 7.10(a) shows again the universality of the spin conductance fluctuations with respect to a wide range of parameters characterizing edge roughness, ribbon length and width. There is little dependence on the ribbon width W , confirming that the observed effect is entirely due to the edges.

Up to now, we have focussed on the case of dilute disorder $d \gg a$. However, currently there is not much experimental control over the edges of nanoribbons. Hence, such nanoribbons may very well exhibit dense disorder, where d is on the order of the lattice constant, $d = \mathcal{O}(a)$. In Fig. 7.10(b) we show the maximum value of spin conductance fluctuations as a function of d/a . We observe that the maximum of the spin conductance fluctuations is independent from the maximum height of the steps s , and retains the maximum universal value rms $G_s = 0.4e/4\pi$ for $d \gtrsim 5a$. For $d < 3a - 5a$, the maximum value decreases rapidly. The maximum of the spin conductance fluctuations starts to decrease, when the system remains always in the strongly localized regime for the energy range corresponding to the edge state, since the conductance fluctuations are suppressed in this case. Still, even in this regime, rare events will lead to spin-polarized currents, as discussed in Section 7.3.2. For $d \gtrsim 5a$, the edge state still has an appreciable transmission probability, and therefore we observe the universal maximum of the spin conductance fluctuations. Again, this can be understood from

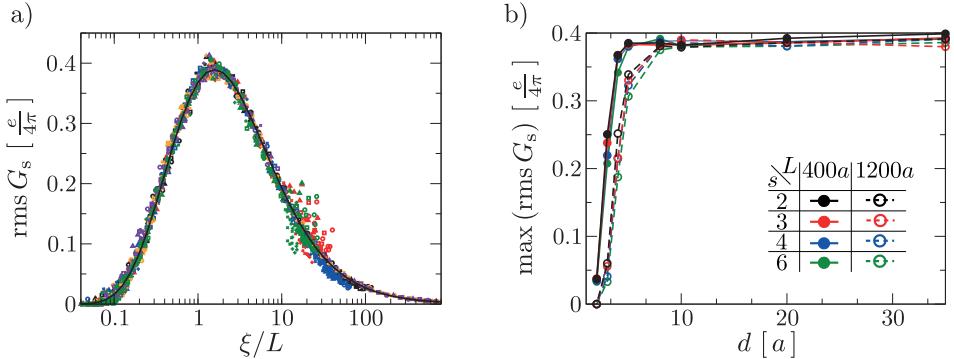


Figure 7.10 – (a) rms G_s as a function of ξ/L for step disorder in n-doped graphene: $d = 20a$ and $s = 3$ (red; orange for $W = 94a/\sqrt{3}$), $d = 35a$ and $s = 2$ (black), $d = 35a$ and $s = 6$ (blue; violet for $W = 94a/\sqrt{3}$), $d = 20a$ and $s = 6$ (green). The solid line corresponds to the DMPK prediction, Eq. (7.14). The data is shown for GNR lengths $L = 800a$ (\circlearrowleft), $1000a$ (\square), $1200a$ (\triangle), $1400a$ ($+$), and $1600a$ (\times), width $W = 34/\sqrt{3}a$ unless specified otherwise. The rms G_s is estimated from 1000 ($W = 32a/\sqrt{3}$) and 750 ($W = 92a/\sqrt{3}$) disorder configurations. (b) Maximum value of rms G_s as a function of d/a for the step disorder models.

the fact that a few (3–4) hexagons in zigzag orientation already support a local edge state [169]. Hence, the transport properties of the edge state are surprisingly stable with respect to edge disorder, just as the magnetic properties of the edge state at disordered edges [195, 202–205].

7.3.4 All-electrical detection of edge magnetism

Although there is evidence for the existence of the edge state at graphene edges on graphite through STM measurements [172–174], neither the existence of the edge state in exfoliated graphene nor the edge magnetism have been proven experimentally yet. In fact, both STM on exfoliated graphene edges and magnetic STM pose many experimental challenges [212]. Here, we propose an experimental alternative to STM for proving the existence of edge magnetism using an all-electrical measurement only. This measurement makes use of the finite spin conductance G_s of a zigzag nanoribbon, predicted in the previous sections, and is based on the ideas put forward in Ref. [213].

Since the edge magnetization breaks time-reversal symmetry, the finite spin conductance G_s can be in principle measured both by two- or four probe setups involving ferromagnetic leads [214]. Fig. 7.11 shows a possible non-local four-probe measurement setup, similar to the experimental setups for spin injection into bulk graphene [43–46]: Current is driven through a ferromagnetic contact, injecting spins into bulk graphene. The spin-polarized charge current from the ferromagnet leads to a spin accumulation underneath the ferromagnetic contact. This spin accumulation can be understood as a

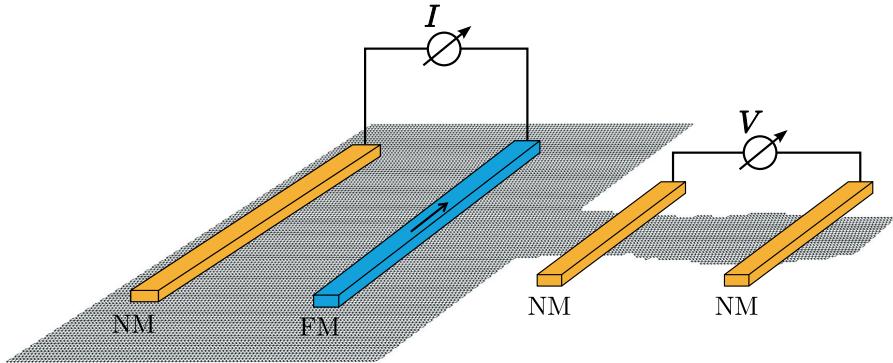


Figure 7.11 – Possible four-probe measurement setup for an all-electrical detection of edge magnetism: Spin polarized current is injected via a ferromagnetic (FM) contact (shown in blue), creating a spin accumulation. The diffusing spin accumulation causes a voltage drop due to the finite spin conductance G_s of the nanoribbon that can be measured by non-magnetic (NM) contacts (shown in orange).

difference in chemical potential for spin up and down. Hence, such a spin accumulation leads to a pure spin current. In a nonmagnetic material, this spin current does not lead to a voltage drop, as both spin up and down are conducted equally. However, in the graphene nanoribbon with $G_s \neq 0$ the conductances for spin up and down are different. Hence, the pure spin current would be converted into a charge current, if the circuit were closed, but since no current may flow through the (unbiased) graphene nanoribbon, this leads to a voltage drop along the nanoribbon. This voltage drop can be measured with non-magnetic contacts. Moreover, the measured voltage depends on the orientation of the injected spins. Note that these mechanisms are explained in more detail for the case of diffusive systems in Ref. [215], and for the case of coherent transport in Ref. [213].

A measured voltage drop that depends on the orientation of the magnetization in the ferromagnet would be an unambiguous sign of a finite spin conductance. The existence of a finite spin conductance would then in turn also be evidence for edge magnetism.

7.4 Edge-state induced Spin Hall effect

7.4.1 Spin-dependent deflection of the edge state

We now come back to the spin-dependent deflection of the edge states already observed in Fig. 7.3, and shown again in Fig. 7.12(a). Here, a zigzag graphene nanoribbon with edge magnetization is coupled to a larger (non-magnetic) ribbon with a negative potential V_{gate} , in order to simulate doped bulk graphene with many channels.

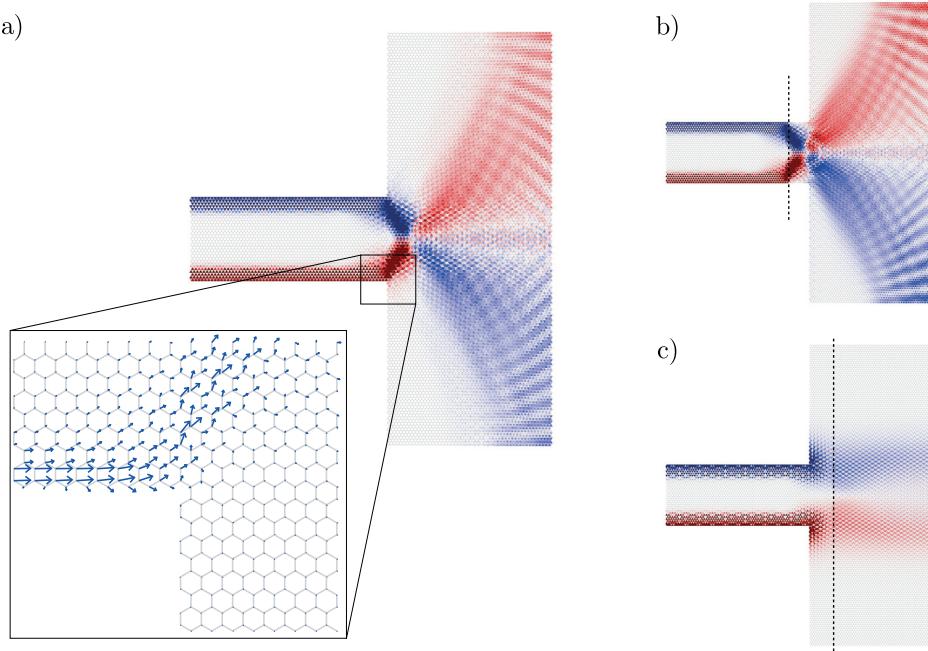


Figure 7.12 – Spin injection from a perfect zigzag GNR into a wider ribbon. Spin up (down) non-equilibrium densities are shown in red (blue). In order to simulate spin injection into bulk graphene, where many transport channels are available, the band edge of the wider graphene ribbon has been shifted by a potential $V_{\text{gate}} = -0.6t$, mimicking the effect of a gate voltage. This potential is switched on (a) directly at the interface between the narrow and the wide ribbon, (b) some distance within the narrow ribbon, and (c) some distance within the wide ribbon (as indicated by black dashed lines). The inset in (a) shows the current density close to one of the edges.

When the edge state enters the bulk, it must expand, as the solution strongly localized to the zigzag edge is not a bulk solution. Furthermore, because of the edge geometry at the contact between the narrow and the wide ribbon, one might expect quantum mechanical diffraction at the edge. At first glance we might thus expect a uniform expansion of the edge state, possibly enhanced in the direction of the edge due to diffraction. This is *not* observed in the numerical simulations: Instead, we see a clear deflection *away* from the graphene edge. Since spin up and down are located on opposite edges, they are deflected in opposite directions and hence separated, leading to a *spin Hall effect*, similar to systems with spin-orbit interaction [25–28].

In order to understand this effect, first it should be noted that even if the geometry

of the system is symmetric with respect to the ribbon axis, as in Fig. 7.12(a), the staggered magnetization breaks this symmetry. The Hamiltonian of the system is then only invariant under reflection about the ribbon axis and $s^z \rightarrow -s^z$, i.e. exchange of spin up and spin down. This can be seen from Eq. (7.2) and the fact that a reflection about the ribbon axis interchanges sublattice A and B. Hence, there is *no* symmetry that would *exclude* a spin Hall effect. Moreover, being localized towards one zigzag boundary, the edge state experiences a highly asymmetric scattering situation.

Second, by closer inspection we observe that the deflection already starts *before* the narrow GNR opens into the bulk. This is particularly obvious from the behavior of the current density close to the edge, as shown in the inset of Fig. 7.12(a), and is due to evanescent waves, that tend to be more important in graphene than in usual semiconductor electron gases (for an example see Ref. [216]; a further example are the (spin) currents in mesoscopic Hall bridges presented in the next section). Furthermore, from the current density we also observe that the edge state starts to expand uniformly (there is both current flowing towards the edge and away from the edge). However, the part expanding towards the edge is immediately reflected, leading to an overall deflection away from the edge.

Hence, it seems that the deflection is simply a geometrical effect due to the localized nature of the edge state. This picture is supported by simulations, where we changed the position where the potential V_{gate} is switched on: In Fig. 7.12(b) the potential is switched on within already the narrow ribbon, and we observe again a deflection away from the edge. In contrast, when the potential is switched on after some distance within the wide ribbon, the edge state expands uniformly, as there is *no edge* at the point where the state starts to expand. Furthermore, numerical simulations show that even without a magnetization, the (then unpolarized) edge state deflects away from the edge, ruling out the valley effects discussed in Ref. [217].

In order to demonstrate that this deflection also leads to a spin-polarized current, we consider a three-terminal T-bar structure, as shown in the insets of Fig. 7.13(a) and (b), where we consider transport from lead 1 to the leads 2 and 3. Unlike before, we do not apply an additional potential V_{gate} .

The spin conductance from lead l' to lead l is then defined as

$$G_{ll'}^s = G_{ll'}^\uparrow - G_{ll'}^\downarrow \quad (7.15)$$

where $G_{mn}^{\uparrow(\downarrow)}$ are the corresponding charge conductances for spin up (down). Note that due to the symmetry reasons discussed above, $G_{21}^{\uparrow(\downarrow)} = G_{31}^{\downarrow(\uparrow)}$, and hence $G_{21}^s = -G_{31}^s$.

In Fig. 7.13(a) we plot G_{21}^s as a function of energy for a geometry where a narrow zigzag GNR contacts a very wide armchair ribbon. This situation is close to a zigzag GNR contacting bulk graphene. In the n-type regime, we observe an on average positive spin conductance, in agreement with the previous spin density pictures (Fig. 7.12). The spin conductance exhibits sharp peaks corresponding to energies, where a new channel opens in the wide nanoribbon; this is remanence of the still finite width. In the p-type regime, the spin conductance is negative, because spin up and down are then located on the opposite edge as compared to the n-type case.

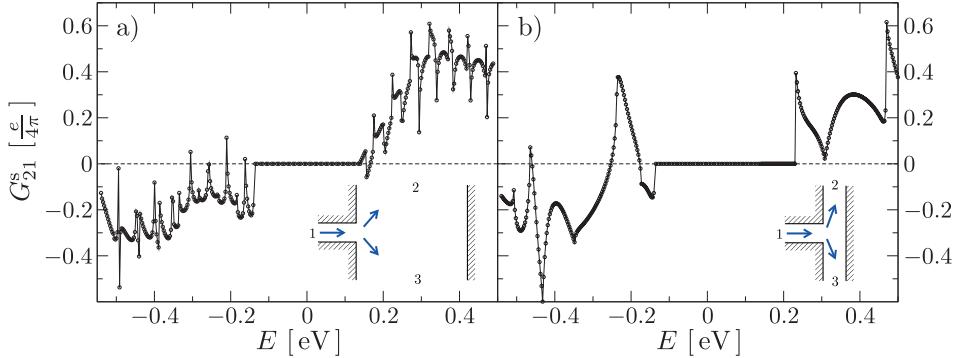


Figure 7.13 – Spin Hall effect: G_{21}^s as a function of energy E in a T-bar geometry with a zigzag GNR (lead 1) with width $W_1 = 75.5a/\sqrt{3} \approx 36a$ and (a) an armchair GNR (lead 2 and 3) with width $W_2 = W_3 = 152a$ and (b) an armchair GNR with width $W_2 = W_3 = 31a$.

In contrast to the case, where we inject spins into a wide armchair ribbon, Fig. 7.13(b) shows the spin conductance G_{21}^s for a geometry consisting of a narrow zigzag ribbon and an armchair ribbon of comparable width. In this case, the number of channels in the zigzag and armchair ribbon are comparable, and the arguments from above (in particular, the “expansion” of the edge state, based on bulk graphene with many channels) only apply to some extent here. However, the scattering situation is again asymmetric for an edge state localized to a boundary. Correspondingly, we again observe a finite spin Hall effect, but the spin conductance G_{21}^s exhibits larger fluctuations related to the number of open channels in the armchair ribbon, and also changes in sign. Still, the overall trend of the sign of G_{21}^s is compatible with the deflection picture.

7.4.2 Pure spin currents in graphene micro-bridges

After using the spin-dependent deflection of the edge state in order to generate spin-polarized currents in a T-bar geometry, we now turn to the problem of generating pure spin currents, i.e. spin currents without corresponding charge currents.

For this, we use a mesoscopic Hall-bar geometry as sketched in Fig. 7.14. Current is flowing through the leads 1 and 4, whereas the leads 2 and 3 act as voltage probes, such that $I_2 = I_3 = 0$. We choose lead 4 as the voltage reference, and set $V_4 = 0$. In linear response, the currents I_l and the voltages V_l are related through Eq. (2.118)

$$I_l = \sum_{l'} G_{ll'} (V_{l'} - V_l), \quad (7.16)$$

and for given values of the currents I_l , the voltages V_l are determined uniquely (see, for example, Ref. [59, chap. 4]). Whereas the charge currents $I_l = I_l^\uparrow + I_l^\downarrow$ are strictly

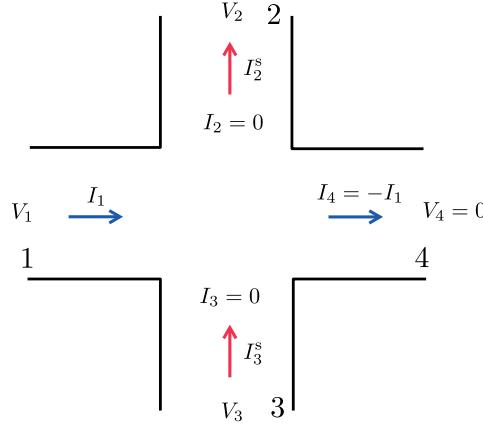


Figure 7.14 – Hall-bar geometry used to generate pure spin currents. Leads 1 and 4 are made of zigzag GNRs, whereas leads 2 and 3 are armchair GNRs.

zero within the voltage probes, the *spin currents*

$$I_l^s = \frac{\hbar}{2e} (I_l^\uparrow - I_l^\downarrow) \quad (7.17)$$

can be very well finite. Here, $I_l^{\uparrow(\downarrow)}$ is defined as the current carried by spin up (down). Note that for the system under consideration the definition of the spin-current is not plagued by ambiguities, in contrast to systems with spin-orbit coupling [218], as spin is a good quantum number for a collinear magnetization. The *spin Hall conductance* of the micro-bridge is defined as [219]

$$G_{\text{sH}} = \frac{I_2^s}{V_1 - V_4} = \frac{I_2^s}{V_1}, \quad (7.18)$$

i.e. as the spin current response to a applied voltage. Since spin up and down form two independent subsystems, the spin currents are simply given as

$$I_l^s = \sum_{l'} G_{ll'}^s (V_{l'} - V_l), \quad (7.19)$$

where $G_{ll'}^s$ is defined as in Eq. (7.15). Hence, the spin Hall conductance, Eq. (7.18), is given as

$$G_{\text{sH}} = G_{21}^s + G_{23}^s \frac{V_3}{V_1} - (G_{21}^s + G_{23}^s + G_{24}^s) \frac{V_2}{V_1}. \quad (7.20)$$

Before evaluating G_{sH} numerically, it is useful to discuss its behavior for symmetric systems. Consider a Hall-bar that has a reflection symmetry about the y -axis, such as

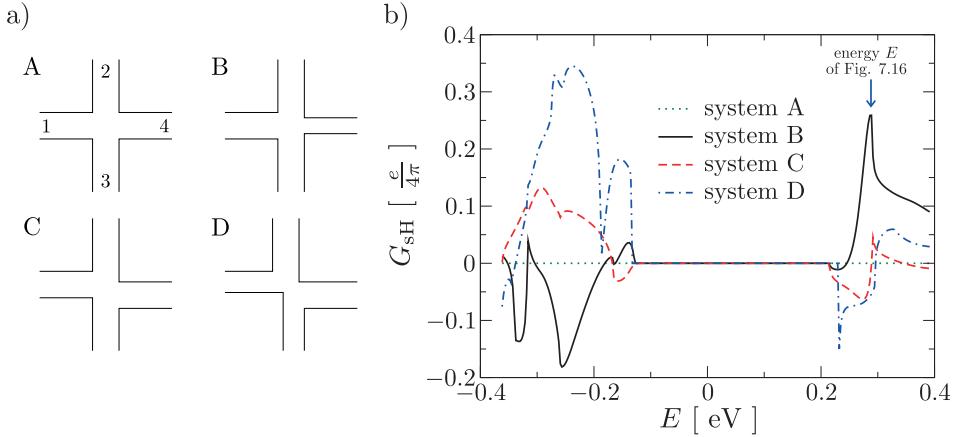


Figure 7.15 – (a) Different geometries considered for generating pure spin currents: system A is symmetric with respect to the axis spanned by leads 2 and 3, whereas systems B, C, and D have been slightly desymmetrized. A: $W_1 = W_4 = 49a/\sqrt{3} \approx 28a$, $W_2 = W_3 = 32a$. B: as system A, but $W_4 = 46a/\sqrt{3}$. C: as system A but lead 4 is shifted by $2a$. D: as in system A, but all leads are shifted and varied in width by $1 - 4a$. Note that for clarity of the presentation, the asymmetries are exaggerated in the picture. (b) Spin Hall conductance G_{sH} as a function of energy E for the Hall-bar geometries shown in (a).

the geometry shown in Fig. (7.14). This reflection maps lead 1 into lead 4 and vice versa, but keeps leads 2 and 3 unchanged. Then, $G_{24}^{\uparrow(\downarrow)} = G_{21}^{\uparrow(\downarrow)}$ and hence $G_{24}^s = G_{21}^s$. Eq. (7.20) then simplifies to

$$G_{sH} = G_{21}^s \left(1 - \frac{2V_2}{V_1} \right) + G_{23}^s \left(\frac{V_3}{V_1} - \frac{V_2}{V_1} \right). \quad (7.21)$$

Moreover, the reflection symmetry about the y -axis leads to $V_2 = V_3 = V_1/2$ [219], and therefore in a perfectly symmetric Hall-bar $G_{sH} = 0$. In such a symmetric system it is thus not possible to generate pure spin currents, but only spin-polarized currents.

This is fundamentally different from a system with spin-orbit coupling, where the Hamiltonian of the system is only invariant under a reflection and a spin flip, leading to $G_{24}^{\uparrow(\downarrow)} = G_{21}^{\uparrow(\downarrow)}$ (and, in addition, spin-flip terms) and a finite spin Hall conductance even for symmetric geometries, as seen for the case of Rashba spin-orbit coupling in Ref. [219]. In the graphene Hall-bar, spin is a good quantum number, and the spin Hall effect arises only due to the geometry, i.e. the boundaries of the system. In particular, the spin Hall conductance vanishes, when the hall voltages V_2 and V_3 of the full system (spin up and down) also lead to $I_2^\uparrow = I_2^\downarrow = 0$, i.e. if the spin subsystems exhibit the same hall voltages as the full system. This is the case for a symmetric Hall-bar. Nevertheless, already little deviations from perfect symmetry are enough to obtain a finite spin Hall

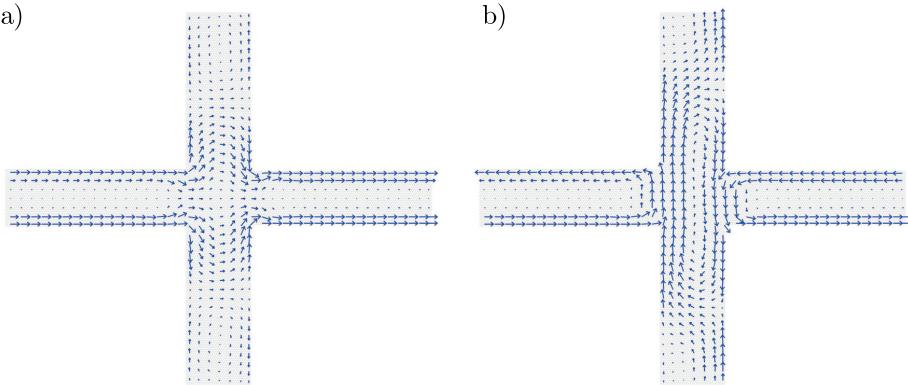


Figure 7.16 – (a) Charge and (b) spin current density in a desymmetrized Hall-bar, corresponding to system B in Fig. 7.15(a). The energy $E \approx 0.285$ eV, corresponding to the peak indicated with an arrow in Fig. 7.15(b). In both (a) and (b), the current densities have been averaged over several hexagons in order to bring out the characteristics of the current flow.

conductance, as shown below.

Fig. 7.15(b) shows the spin Hall conductance G_{SH} as function of energy E for a symmetric Hall-bar (system A in Fig. 7.15(a)) and slightly desymmetrized Hall-bars (system B, C and D in Fig. 7.15(a))). Note that the asymmetry in the picture is exaggerated to show the difference in the systems more clearly. The real asymmetry involves only a few hexagons, as seen from the numbers in the figure caption.

As expected from the symmetry considerations above, the spin Hall conductance of the symmetric Hall bar vanishes. Nevertheless, the slightly desymmetrized Hall-bars show a finite spin Hall conductance $G_{\text{SH}} = 0.1 - 0.4 \frac{e}{4\pi}$. Thus, the spin Hall conductance has a magnitude similar to the mesoscopic spin Hall effect in semiconductor-based two-dimensional electron gases with spin-orbit coupling [219, 220]. Furthermore, we observe that the spin Hall conductance depends strongly on the geometrical details of the Hall-bar.

Finally, in Fig. 7.16 we show an example of both the charge (Fig. 7.16(a)) as well as the spin current density (Fig. 7.16(b)) in a desymmetrized Hall-bar. Whereas no net charge current is flowing in the voltage probes, we observe that the charge current from the current probes extends rather far into the voltage probes, and in addition, we observe current vortices within the voltage probes. These are due to evanescent modes, that tend to be of great importance in graphene (see, for example, Ref. [216]). In the zigzag ribbons, the spin current is flowing along the edges, and in opposite directions for opposite edges. This is due to the fact that opposite edges carry opposite spins. In addition, we also observe a finite, pure spin current flowing in the transverse direction, as expected from a finite G_{SH} .

7.5 Summary

In this chapter we have discussed various possibilities to generate spin-polarized charge and pure spin currents from the edge magnetism in zigzag graphene nanoribbons. These predictions were verified by numerical simulations for a mean-field model of the edge magnetism within the next-nearest neighbor approximation.

In particular, we have found that edge disorder in zigzag GNRs naturally leads to spin-polarized currents, as it breaks the symmetry between the two oppositely spin-polarized edges. If a zigzag nanoribbon is fabricated such that one of the edges is more clean than the other edge, conductance will be preferentially carried by the spin-polarized edge state at the cleaner edge, leading to a finite spin conductance. Even if both edges are equally rough, single zigzag GNRs exhibit a finite spin conductance due to spin conductance fluctuations. These spin conductance fluctuations exhibit a universal maximum value of rms $G_s \approx 0.4 \frac{e}{4\pi}$ and are independent of the type of disorder at the edges, and also independent of the details in the mean-field model for the edge magnetism.

In addition, the spin-polarized edge states can give rise to a spin Hall effect. This spin Hall effect can be used to generate spin-polarized currents in three-terminal geometries, and pure spin currents in four-terminal geometries. In contrast to the spin-polarized currents in rough graphene nanoribbons, the spin Hall effect requires ballistic systems, and does not exhibit a universal behavior, as the effect is due to the details of the underlying geometry.

The results of this chapter show that zigzag nanoribbons are an efficient alternative to ferromagnetic metals for generating spin-polarized and pure spin currents, thus paving the way to an all-graphene spintronics.

Chapter 8

Summary and perspectives

8.1 Summary

In the first part of this thesis we have developed algorithms for a generic approach to quantum transport calculations, valid for arbitrary tight-binding models.

From general considerations within the non-equilibrium Green's function theory we have found that for such a generic approach it is necessary to solve two problems: first, the calculation of the surface Green's function of a lead and second, the efficient calculation of the retarded Green's function of the scattering region. In order to be generally applicable, these problems had to be solved without restrictive assumptions on the underlying tight-binding model or the geometry of the system.

The problem of calculating the surface Green's function of the leads was approached in Chapter 3. The lead was modeled as a semi-infinite wire consisting of a periodic repetition of unit cells. Under that premise, we derived an expression for the lead Green's function in terms of Bloch eigenstates of the corresponding infinite wire, extending previously known results. A direct implementation of this expression was found to be numerically unstable in certain, important examples. Motivated by insights from an analysis of this failure, we developed a numerically stable algorithm to calculate the lead Green's function in terms of the Schur decomposition of a certain matrix, in contrast to the direct implementation that was based on the eigendecomposition.

Chapter 4 was then concerned with the efficient calculation of the retarded Green's function. To this end we analyzed several well-established quantum transport algorithms, such as the recursive Green's function method. We found that a wide class of transport algorithms depends on the Hamiltonian matrix being in a block-tridiagonal form; this is the only essential requirement. Based on graph partitioning techniques we developed a matrix reordering technique to bring an arbitrary matrix into a block-tridiagonal form suitable and optimal for transport. Through this reordering, well-

established quantum transport algorithms, that were restricted to wire geometries in their original formulation, can be applied to any tight-binding model and any geometry, including multi-terminal systems. In addition we found that matrix reordering can also lead to a significant speed-up of numerical calculations. These advantages of using the matrix reordering technique were then demonstrated on representative examples.

The results of Chapters 3 and 4, together with the generalized Fisher-Lee relation of Appendix C, form the foundation of a generic computer program, applicable to any tight-binding system. In order to adapt this approach to a certain tight-binding model, the only remaining task is to set up the respective Hamiltonian matrix. This may still be laborious, but is straightforward. For the purpose of this work, we considered the tight-binding description arising from the finite differences approximation to the effective mass Schrödinger equation in two and three dimensions (Appendix F), as well as the next-nearest neighbor tight-binding model for graphene (Appendix G).

In the second part of this thesis we then applied the numerical methods developed in the first part to study spin-dependent effects in graphene and magnetic tunnel junctions.

The tunneling anisotropic magnetoresistance effect in an epitaxial Fe/GaAs/Au tunnel junction was the subject of Chapter 5. In particular, we investigated the magnetic field dependence of the TAMR effect in order to explain recent experimental findings. Previous work suggested that the origin of the TAMR effect in this magnetic tunnel junction is due to spin-orbit coupling in the GaAs barrier, and presented a corresponding microscopic model. In this work, we found that it is enough to include the orbital effects of a magnetic field into this model to explain the experimental observations. Both numerical results and a phenomenological model reproduce all characteristic features of the measured magnetic field dependence of the TAMR effect. This magnetic field dependence could be explained by an interplay of orbital effects due to the magnetic field and the Dresselhaus spin-orbit coupling within the GaAs barrier.

Finally, Chapters 6 and 7 were concerned with charge and spin transport properties of the graphene edge state in zigzag graphene nanoribbons.

The results of previous work seemed to indicate very surprising transport properties of the edge state: Although the probability density is located close to the edges of the ribbon, current flows through the middle of the ribbon, rendering the edge state rather insensitive to edge disorder. In Chapter 6, we have found these counter-intuitive results to be an artifact of the nearest-neighbor tight-binding model used almost exclusively for transport calculations in graphene. In fact, exponentially small corrections to this model in the form of next-nearest neighbor hopping or a position-dependent potential or staggered potential were found to change the transport properties fundamentally, restoring the current flow to the edge, in accordance with the probability density. This then also leads to a more intuitive transport behavior: The edge state is strongly scattered by defects located on the same edge, whereas it remains virtually unaffected by defects located on the opposite edge. These effects could be understood by symmetry considerations within perturbation theory. From these findings we have concluded that

nearest-neighbor tight-binding does not apply to any realistic system where the edge state plays a role, and hence strongly advocate the inclusion of next-nearest neighbor hopping.

Based on the theoretically predicted edge magnetism in zigzag graphene nanoribbons, we have proposed several schemes for generating spin-polarized and pure spin currents within graphene nanoribbons in Chapter 7. For this we have employed a mean-field model of the edge magnetism within the next-nearest neighbor tight-binding approximation of graphene, which, as seen from Chapter 6, is suitable for studying edge state properties. We have found that edge disorder in zigzag graphene nanoribbons naturally leads to spin-polarized currents. In particular, rough graphene nanoribbons exhibit spin conductance fluctuations with a universal maximum value of rms $G_s \approx 0.4 e/4\pi$, independent of the type of edge disorder and the details of the mean-field model for the edge magnetism. Based on the finite spin conductance of zigzag nanoribbons, we have suggested an all-electrical measurement to prove the existence of the edge magnetism that has been experimentally elusive so far. Moreover, we have found that the edge state can lead to a geometry-induced spin Hall effect in ballistic nanostructures. This spin Hall effect can be used to generate spin-polarized currents in three-terminal devices, and pure spin currents in four-terminal structures. These results show that zigzag nanoribbons are an efficient alternative to ferromagnetic metals for generating spin currents, paving the way to an all-graphene based spintronics.

8.2 Outlook

We have emphasized that the generic framework for numerical transport calculations developed in this thesis is applicable to arbitrary tight-binding systems, beyond the examples discussed in this work. For example, the code has also been used to study the topological Hall effect [57]. Moreover, Viktor Krückl [56] has extended the existing program code to study transport of heavy holes within the Luttinger-Kohn Hamiltonian [221]. Applying this transport framework to other systems is just as well possible, including tight-binding models originating from density functional theory calculations, as in Refs. [84–86].

Note that the discussion in this work was restricted to orthogonal tight-binding models. The extension to nonorthogonal tight-binding models is straightforward, but tedious, as the use of a nonorthogonal basis involves some subtleties: For example, the anticommutation relations for creation and annihilation operators c_i , c_j^\dagger then involve the overlap matrix S instead of a Kronecker delta (for example, see Ref. [222]):

$$\{c_i, c_j^\dagger\} = \delta_{ij} \quad \rightarrow \quad \{c_i, c_j^\dagger\} = (S^{-1})_{ij}. \quad (8.1)$$

Thus, extending the theory to nonorthogonal tight-binding models requires some care.

The theoretical and experimental work on the TAMR effect in Fe/GaAs/Au tunnel junctions has always involved mutual stimulations between theorists and experimental-

ists here in Regensburg, both in the original work [47] as well as in the work presented here. From the model presented in Chapter 5 we have made some predictions on the magnetic field behavior of the conductance that can be tested experimentally. In turn, new experimental data may lead to further extensions of the existing theory.

The field of graphene physics has seen a tremendous evolution in the past four years and still contains many unsolved puzzles. In this work, we were concerned only with one aspect of graphene, namely the graphene edge state and its spin transport properties.

In this study, the role of the edges was of essential importance. In fact, edges are expected to influence many aspects of graphene: For example, there have been several studies on the influence of edge disorder on the (charge) conductance of graphene nanoribbons [162, 165, 166] in order to explain the yet ununderstood transport gap of graphene nanoribbons [158–160]. All of these studies are carried out using the nearest-neighbor tight-binding model for graphene. As seen discussed in Chapter 6, this model leads to spurious transport results for the edge state. Especially in the low-energy regime, where the transport gap is experimentally observed, the edge state is expected to play an important role. Hence, the transport calculations in Refs. [162, 165, 166] should be tested against a model including next-nearest neighbors.

Edges have already been shown to be of crucial importance for the symmetry classes in ballistic graphene cavities [50], and may be of relevance in recent experiments on graphene quantum dots [223]. On the other hand, in diffusive graphene the chiral nature of the charge carriers is expected to show in a weak antilocalization effect, if intervalley scattering is suppressed [224]. However, the weak antilocalization observed in the experiment has often been more or less suppressed with respect to the theoretically predicted value (for examples, see Refs. [225–227]). Again, the edges may play an important role in this case. Obviously in narrow graphene nanoribbons, intervalley scattering from sharp edge disorder will dominate, suppressing weak antilocalization [224]. However, also for samples that have a width comparable to their length the influence of the edges may be significant [228]. These predictions can be conveniently tested by numerical simulations.

Interest in graphene has grown tremendously in the past few years, and will continue to yield fascinating physics in the future. Here, we have only mentioned a few examples, that are related to the topics discussed in this work. Especially the numerical techniques developed in the course of this thesis will be of great use for future studies on graphene.

Appendix A

Observables in tight-binding

A.1 Single-particle operators in tight-binding

Expressions for observables in envelope approximation, such as charge or current densities, can be derived quite generally for any tight-binding Hamiltonian. For that purpose, we consider a tight-binding Hamiltonian including spin with orbitals $\chi_{i,s}(\mathbf{x})$, where i is the index of the lattice site \mathbf{x}_i and $s = 1, 2$ the spin index. Without loss of generality, the discussion is restricted to a single orbital per lattice site, and the orbitals are assumed to be eigenstates of σ^z . The Hamiltonian is given by

$$H = \sum_{i,j} \sum_{s,s'=1}^2 t_{ij,ss'} |is\rangle \langle js'| . \quad (\text{A.1})$$

Since H must be a hermitian operator, $t_{ij,ss'} = t_{ji,s's}^*$.

The density operator in tight-binding representation is given by

$$\begin{aligned} n(\mathbf{x}) &= \sum_{i,j,s,s'} \int d\mathbf{x}' \chi_{i,s}(\mathbf{x}')^* \delta(\mathbf{x} - \mathbf{x}') \chi_{j,s'}(\mathbf{x}') |is\rangle \langle js'| \\ &= \sum_{i,j,s} \chi_{i,s}(\mathbf{x})^* \chi_{j,s}(\mathbf{x}) |is\rangle \langle js| . \end{aligned} \quad (\text{A.2})$$

This expression contains an overlap between orbitals of different lattice sites. If the orbitals are sufficiently localized, this overlap can be neglected and only the orbital i closest to \mathbf{x} will contribute to the density:

$$n(\mathbf{x}) = \sum_s |\chi_{i,s}(\mathbf{x})|^2 |is\rangle \langle is| . \quad (\text{A.3})$$

Furthermore, if we are only interested in the variation of the wave function's *envelope*, we can neglect the variations of the orbitals. For that purpose we approximate the

atomic orbitals by the characteristic function of a volume (or area) V around the lattice site \mathbf{x}_i :

$$\chi_{i,s}(\mathbf{x}) = \begin{cases} \frac{1}{V} & \text{if } \mathbf{x} \text{ is in the vicinity of } \mathbf{x}_i, \\ 0 & \text{else.} \end{cases} \quad (\text{A.4})$$

To keep the notation simple, we assume that V is the same for each lattice point¹. With that, we arrive at the final expression for the electron density operator

$$n(\mathbf{x}_i) = \frac{1}{V} \sum_s |is\rangle \langle is|, \quad (\text{A.5})$$

where \mathbf{x}_i should be understood to represent any point \mathbf{x} in the vicinity of \mathbf{x}_i . In a similar manner we can derive the tight-binding representation of the spin density operator

$$n^\mu(\mathbf{x}_i) = \frac{1}{V} \sum_{s,s'} \sigma_{ss'}^\mu |is\rangle \langle is'| \quad \text{for } \mu = 0, x, y, z, \quad (\text{A.6})$$

where $\sigma^0 = \mathbb{1}$, so that Eq. A.6 also contains the previously derived $n(\mathbf{x}_i)$.

For the current density, it is necessary to derive an expression for the velocity operator first. The position operator is given by

$$\begin{aligned} \mathbf{x} &= \sum_{i,j,s,s'} \int d\mathbf{x}' \chi_{i,s}(\mathbf{x}')^* \mathbf{x}' \chi_{j,s'}(\mathbf{x}') |is\rangle \langle js'| \\ &= \sum_{i,s} \mathbf{x}_i |is\rangle \langle is|, \end{aligned} \quad (\text{A.7})$$

where we assume again that the atomic orbitals are sufficiently localized. The velocity operator is related to the time derivative of the position operator:

$$\begin{aligned} \mathbf{v} &= \frac{d}{dt} \mathbf{x} = \frac{i}{\hbar} [H, \mathbf{x}] \\ &= \frac{i}{\hbar} \sum_{i,j,s,s'} \sum_{n,t} t_{ij,ss'} \mathbf{x}_n [|is\rangle \langle js'|, |nt\rangle \langle nt|] \\ &= \frac{i}{\hbar} \sum_{i,j,s,s'} (\mathbf{x}_j - \mathbf{x}_i) t_{ij,ss'} |is\rangle \langle js'|. \end{aligned} \quad (\text{A.8})$$

From the density and the velocity operator, we can define the current operator as $\mathbf{j}(\mathbf{x}) = \frac{1}{2} \{n(\mathbf{x}), \mathbf{v}\}$ [229]. Note that in the transition from classical to quantum mechanics, the classical expression for the current, $\mathbf{j} = n\mathbf{v}$, must be symmetrized by the anti-commutator $\{\cdot, \cdot\}$ to obtain a Hermitian operator [153, Chap. 3]. Inserting Eqs. (A.2) and (A.8), we obtain

$$\mathbf{j}(\mathbf{x}_i) = \frac{i}{2\hbar V} \sum_{j,s,s'} (\mathbf{x}_j - \mathbf{x}_i) (t_{ij,ss'} |is\rangle \langle js'| - t_{ji,s's} |js'\rangle \langle is|). \quad (\text{A.9})$$

¹ For example, this is true for any grid based on a Bravais lattice.

When local currents are considered in tight-binding models, this is often done in terms of *bond currents* I_{ij} between two sites i and j [179, 230]. The current density operator (A.9) is equivalent to these bond currents, but contains in addition the distance vector between two lattice points ($\mathbf{x}_j - \mathbf{x}_i$) (leading to the vector character of $\mathbf{j}(\mathbf{x})$, these are usually put in by hand for the bond currents) and a constant prefactor (giving the correct units for a current density).

Analogous to the current density operator, we also define the spin current density operator as $\mathbf{j}^\mu(\mathbf{x}) = \frac{1}{2} \{n^\mu(\mathbf{x}), \mathbf{v}\}$, where again $\mu = 0$ contains the previous expression for the current density, Eq. (A.9). Then we find

$$\mathbf{j}^\mu(\mathbf{x}_i) = \frac{i}{2\hbar V} \sum_{j,s_1,s_2,s_3} (\mathbf{x}_j - \mathbf{x}_i) (\sigma_{s_1 s_3}^\mu t_{ij,s_3 s_2} |is_1\rangle \langle js_2| - \sigma_{s_3 s_1}^\mu t_{ji,s_2 s_3} |js_2\rangle \langle is_1|) . \quad (\text{A.10})$$

Note that there are several inequivalent definitions for the spin current. In addition, if the spin operator does not commute with the Hamiltonian, as e.g. for systems with spin-orbit interaction, these spin currents do not obey current conservation laws [218]. However, in this work we need not worry about such subtleties, as we will consider the spin current only in systems where spin is a good quantum number. In particular, when $t_{ij,ss'} = t_{ij,s} \delta_{ss'}$, the spin current density simplifies to

$$\mathbf{j}^\mu(\mathbf{x}_i) = \frac{i}{2\hbar V} \sum_{j,s,s'} (\mathbf{x}_j - \mathbf{x}_i) (\sigma_{ss'}^\mu t_{ij,s'} |is\rangle \langle js'| - \sigma_{s's}^\mu t_{ji,s'} |js'\rangle \langle is|) . \quad (\text{A.11})$$

A.2 Observables in terms of Green's functions

The transition from a single-particle operator A to the corresponding many-body operator \mathcal{A} is particularly simple in the case of a tight-binding Hamiltonian: Since $\mathcal{A} = \sum_{i,j,s,s'} \langle is| A |js' \rangle c_{is}^\dagger c_{js'}$ (for an introduction to many-body operators see Ref. [70]), the many-body operator is simply obtained by replacing $|is\rangle \rightarrow c_{is}^\dagger$ and $\langle is| \rightarrow c_{is}$. Observables, i.e. expectation values of these operators, are then given in terms of lesser Green's functions at equal times $t' = t$,

$$G^<(js', t, is, t) = G_{js', is}^<(t, t) = \frac{i}{\hbar} \langle c_{is}^\dagger(t) c_{js'}(t) \rangle , \quad (\text{A.12})$$

as defined in Section 2.5.1. In particular, we find the spin density

$$n^\mu(\mathbf{x}_i, t) = -\frac{i\hbar}{V} \sum_{ss'} \sigma_{ss'}^\mu G_{is', is}^<(t, t) \quad (\text{A.13})$$

and the spin current density

$$\begin{aligned} \mathbf{j}^\mu(\mathbf{x}_i, t) &= \frac{1}{2V} \sum_{j,s_1,s_2,s_3} (\mathbf{x}_j - \mathbf{x}_i) (\sigma_{s_1 s_3}^\mu t_{ij,s_3 s_2} G_{js_2, is_1}^<(t, t) - \sigma_{s_3 s_1}^\mu t_{ji,s_2 s_3} G_{is_1, js_2}^<(t, t)) \\ &= \frac{1}{V} \sum_{j,s_1,s_2,s_3} (\mathbf{x}_j - \mathbf{x}_i) \Re \epsilon (\sigma_{s_1 s_3}^\mu t_{ij,s_3 s_2} G_{js_2, is_1}^<(t, t)) . \end{aligned} \quad (\text{A.14})$$

In the expression for the spin current density we made use of the fact that $\sigma_{ss'}^\mu = (\sigma_{s's}^\mu)^*$, $t_{ij,ss'} = t_{ji,s's}^*$ and Eq. (2.22), $G_{is_1,js_2}^<(t,t) = -(G_{js_2,is_1}^<(t,t))^*$.

In steady-state, the Green's functions only depend on time-differences, and hence the observables are time-independent, as discussed in Chapter 2. In this case it is advantageous to work in the energy domain: From the Fourier transformation of the lesser Green's function, Eq. (2.24), we then find

$$n^\mu(\mathbf{x}_i) = \int dE n^\mu(\mathbf{x}_i, E), \quad (\text{A.15})$$

where

$$n^\mu(\mathbf{x}_i, E) = -\frac{1}{2\pi V} \sum_{ss'} \sigma_{ss'}^\mu G_{is',is}^<(E), \quad (\text{A.16})$$

and

$$\mathbf{j}^\mu(\mathbf{x}_i) = \int dE \mathbf{j}^\mu(\mathbf{x}_i, E), \quad (\text{A.17})$$

where

$$\mathbf{j}^\mu(\mathbf{x}_i, E) = \frac{1}{2\pi\hbar V} \sum_{j,s_1,s_2,s_3} (\mathbf{x}_j - \mathbf{x}_i) \Re e \left(\sigma_{s_1 s_3}^\mu t_{ij,s_3 s_2} G_{js_2,is_1}^<(E) \right). \quad (\text{A.18})$$

The lesser Green's function necessary for the evaluation of these observables are conveniently obtained from the Keldysh equation (2.112).

Note that the energy-dependent observables $n^\mu(\mathbf{x}_i, E)$ and $\mathbf{j}^\mu(\mathbf{x}_i, E)$ are the expectation values of the respective operators in terms of eigenstates of the system at energy E , weighted with the density of states, as seen from Eq. (2.123). As such, they are not directly observable in an experiment.

In linear response however, we can interpret, for example, $n(\mathbf{x}_i, E)$ as a non-equilibrium density. To this end, consider a system with two leads, left (L) and right (R), with $\mu_L > \mu_R$ as in Fig. 2.3. As discussed in Section 2.5.2, in linear response the electrochemical potentials of the leads are close to the Fermi energy, $\mu_R \approx \mu_L \approx E_F$, and the voltage $V = (\mu_L - \mu_R)/e$ is small. We split the density $n(\mathbf{x})$ in two parts,

$$n(\mathbf{x}) = n_{\text{eq}}(\mathbf{x}) + n_{\text{non-eq}}(\mathbf{x}), \quad (\text{A.19})$$

where

$$n_{\text{eq}}(\mathbf{x}) = \int_{-\infty}^{E_F} dE n(\mathbf{x}, E) \quad (\text{A.20})$$

is the equilibrium density, and

$$n_{\text{non-eq}}(\mathbf{x}) = \int_{E_F}^{E_F + eV} n(\mathbf{x}, E) \quad (\text{A.21})$$

is the non-equilibrium density induced by the voltage V . When $n(\mathbf{x}, E)$ does not vary much in the energy set by the voltage, we can then approximate

$$n_{\text{non-eq}}(\mathbf{x}) \approx n(\mathbf{x}, E_F + \frac{eV}{2}) \times eV. \quad (\text{A.22})$$

We can thus interpret the energy-dependent observables evaluated at some energy above E_F as voltage-induced non-equilibrium densities.

Appendix B

The recursive Green's function method

B.1 Basic principles

The recursive Green's function technique (RGF) was pioneered by Thouless and Kirkpatrick [118] and by Lee and Fisher [119] in 1981. A few years later, MacKinnon presented a formulation of the RGF technique that is commonly used today [103]. Since then, the RGF method has been applied to various systems numerous times. There are also a number of extensions, for example to deal with non-equilibrium systems [121, 122].

Instead of calculating the retarded Green's function G_S^r of the scattering region directly from Eq. (2.111), the recursive Green's function method constructs G_S^r by building up the system recursively. To this end, the system is divided into *slices* [103]: In the example of a square grid, lattice points with the same longitudinal coordinate are grouped into vertical slices, as depicted in Fig. B.1. The RGF method starts from a system without any slices in the scattering region, and then successively adds layer by layer, until the full system has been constructed. Obviously, the system can only be partitioned into slices as shown in Fig. B.1, if there are only two, parallel leads, i.e. a geometry similar to a wire. Thus, the RGF method is in its common formulation restricted to geometries with only two collinear leads.

The recursive Green's function method is very often only formulated for the particular case of a square grid. It is however much more generic, as a closer inspection of the prerequisites of the RGF technique shows. In mathematical terms, the RGF method is based on the fact that the tight-binding Hamiltonian H of a two-terminal system can

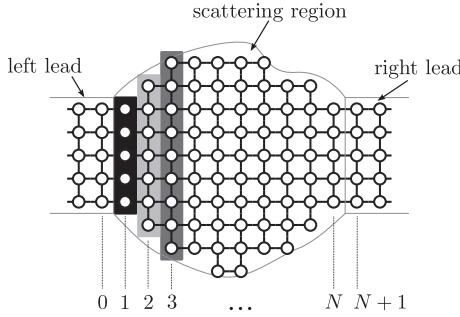


Figure B.1 – A system defined on a square grid can be partitioned into slices i , by grouping lattice points with the same longitudinal coordinate. This kind of partitioning leads to a block-tridiagonal tight-binding matrix, where every slice corresponds to a matrix block. The different slices are shown in alternating shades of grey.

be written in block-tridiagonal form:

$$H = \begin{pmatrix} \ddots & & & & \\ & H_L & V_L & & \\ & V_L^\dagger & H_L & H_{0,1} & \ddots \\ & & H_{1,0} & H_{1,1} & H_{1,2} & 0 \\ & & & H_{2,1} & H_{2,2} & H_{2,3} & \ddots \\ & & & & H_{3,2} & \ddots & & \\ & & & & & \ddots & H_{N-1,N} & \\ & & & & 0 & & H_{N,N-1} & H_{N,N} & H_{N,N+1} \\ & & & & & \ddots & H_{N+1,N} & H_R & V_R \\ & & & & & & V_R^\dagger & H_R & \ddots \end{pmatrix}, \quad (\text{B.1})$$

where the index L (R) denotes the blocks in the left (right) lead, $1 \dots N$ the blocks within the scattering region S, and 0 ($N + 1$) the first block in the left (right) lead. Such a form arises naturally from the partitioning of a square grid as in Fig. B.1, where every slice i corresponds to a matrix block $H_{i,i}$. However, such a block-tridiagonal form can be found for any tight-binding model: In Chapter 4 we developed a matrix reordering algorithm to obtain the optimal block-tridiagonal form for the RGF method for an arbitrary tight-binding Hamiltonian matrix H . This matrix reordering algorithm thus allows the application of the RGF method to any tight-binding system, including multiple, non-collinear leads.

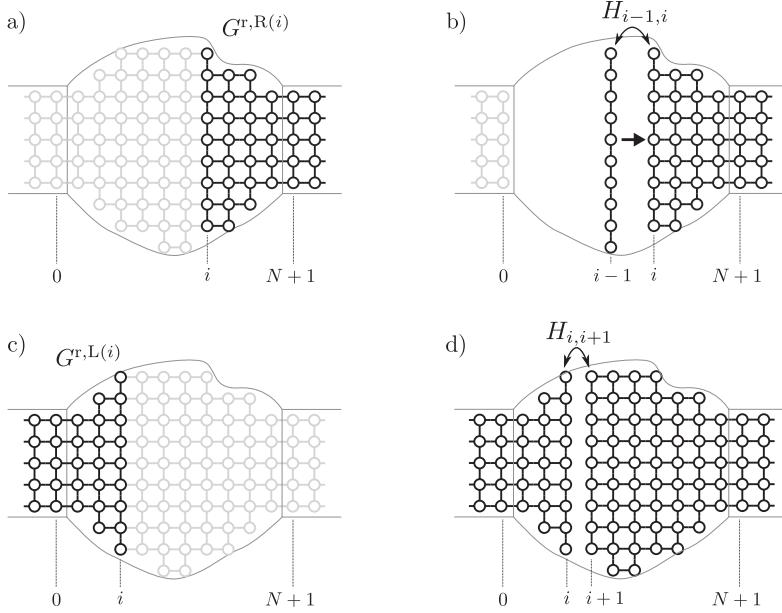


Figure B.2 – Schematic depiction of the recursive Green’s function algorithm: (a) The Green’s function $G^{r,R(i)}$ contains all blocks $\geq i$. (b) $G^{r,R(i-1)}$ can be calculated from $G^{r,R(i-1)}$ by adding block $i-1$ to the structure. (c) The Green’s function $G^{r,L(i)}$ contains all blocks $\leq i$. (d) The full Green’s function of the system is obtained by connecting the still isolated parts described by $G^{r,L(i)}$ and $G^{r,R(i+1)}$.

The central object in the RGF method is the Green’s function $G^{r,R(i)}$ of a system containing only the blocks (slices) with index $\geq i$, as depicted in Fig. B.2(a). Then, there is a relation between $G^{r,R(i)}$ and $G^{r,R(i-1)}$ that can be applied recursively to obtain $G^{r,R(i)}$ for all values of i .

This relation is derived by adding the block $i-1$ to the system, as depicted in Fig. B.2(b). The Green’s functions can then be calculated from the Dyson equation (2.91), $G^r = G_0^r + G_0^r V G^r$. In this situation, $G^r = G^{r,R(i-1)}$, i.e. the Green’s function of the system containing all blocks $\geq i-1$. In contrast, G_0^r is the Green’s function of the system containing all blocks $\geq i$ and the isolated block $i-1$, disconnected from the rest of the system, as described by the Hamiltonian

$$H_0 = \begin{pmatrix} H_{i-1,i-1} & 0 & & \\ 0 & H_{i,i} & H_{i,i+1} & \\ & H_{i+1,i} & H_{i+1,i+1} & H_{i+1,i+2} \\ & H_{i+2,i+1} & & \ddots \end{pmatrix}. \quad (\text{B.2})$$

From this form of the Hamiltonian, we have

$$G_{0,j,k}^r = \begin{cases} G_{j,k}^{r,R(i)} & \text{for } j, k \geq i, \\ (E - H_{i-1,i-1} + i\eta)^{-1} & \text{for } j = k = i-1, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B.3})$$

The isolated block is attached to the system by means of the “perturbation”

$$V = \begin{pmatrix} 0 & H_{i-1,i} & & \\ H_{i,i-1} & 0 & & \\ & & \ddots & \end{pmatrix}. \quad (\text{B.4})$$

Note that V only has matrix elements between the slices $i-1$ and i .

From the Dyson equation, we find, in a calculation similar to that in Section 2.5.2,

$$G_{i-1,i-1}^{r,R(i-1)} = G_{0,i-1,i-1}^r + G_{0,i-1,i-1}^r H_{i-1,i} G_{i,i-1}^{r,R(i-1)} \quad (\text{B.5})$$

and

$$G_{i,i-1}^{r,R(i-1)} = G_{0,i,i}^r H_{i,i-1} G_{i-1,i-1}^{r,R(i-1)}. \quad (\text{B.6})$$

Inserting (B.6) into (B.5) and considering the individual contributions of G_0^r , Eq. (B.3), we arrive at

$$G_{i-1,i-1}^{r,R(i-1)} = \left(E - H_{i-1,i-1} - H_{i-1,i} G_{i,i}^{r,R(i)} H_{i,i-1} \right)^{-1}, \quad (\text{B.7})$$

where we neglected the infinitesimal $i\eta$, as this does not play a role in the numerics, as long as $H_{i-1,i} G_{i,i}^{r,R(i)} H_{i,i-1}$ has a finite imaginary part. Usually this is the case, for instance when there are propagating modes in the structure. In a similar fashion, we also obtain

$$G_{N+1,i-1}^{r,R(i-1)} = G_{N+1,i}^{r,R(i)} H_{i,i-1} G_{i-1,i-1}^{r,R(i-1)} \quad (\text{B.8})$$

and

$$G_{i-1,N+1}^{r,R(i-1)} = G_{i-1,i-1}^{r,R(i-1)} H_{i-1,i} G_{i,N+1}^{r,R(i)}. \quad (\text{B.9})$$

In a similar fashion, we can define the Green's function $G^{r,L(i)}$ that contains all blocks $\leq i$ (see Fig. B.2(c)).

The recursive Green's function calculation then starts from $G_{N+1,N+1}^{r,R(N+1)} = g_R^r$ and $G_{0,0}^{r,L(0)} = g_L^r$, where g_R^r and g_L^r are the surface Green's functions of the right and left lead respectively (on the topic of calculating these surface Green's functions, see Chapter 3). Applying Eqs. (B.7)–(B.9) recursively allows then to calculate $G^{r,R(i)}$ for any i . Note however that, for example, $G^{r,R(1)}$ is not yet the desired Green's function G^r of the full system—it is still disconnect from the left lead. Thus, in order to obtain the full Green's function of the system, the Green's functions $G^{r,R(i+1)}$ and $G^{r,L(i)}$ must be connected via the Dyson equation, as depicted in Fig. B.2(d). Depending on the desired values of the Green's function, this is done in various ways, as summarized in the next section.

In the case of non-interacting systems, or interacting systems in mean-field approximation, the lesser self-energy $\Sigma^<$ only contains contributions from the leads. In this case, the Keldysh equation (2.112) can be written as

$$G_{i,j}^< = G_{i,0}^r \Sigma_L^< (G_{j,0}^r)^\dagger + G_{i,N+1}^r \Sigma_R^< (G_{j,N+1}^r)^\dagger, \quad (\text{B.10})$$

where we made use of $G^a = (G^r)^\dagger$, Eq. (2.21). Here, $\Sigma_L^<$ ($\Sigma_R^<$) is the self-energy contribution from the left (right) lead. It is thus sufficient to compute $G_{i,0}^r$ and $G_{i,N+1}^r$ for $0 \leq i \leq N+1$ in order to obtain all desired values of $G^<$.¹ This is the basic idea behind the extensions of the RGF algorithm to non-equilibrium systems [121, 122].

B.2 Summary of algorithms

Eqs. (B.7) – (B.9) are the building blocks of the various forms of the RGF method. The exact procedure of a calculation depends on which values of the Green's function G^r should be calculated. Here we summarize the computational steps distinguishing a few, prominent examples.

Algorithm B.1. Compute $G_{N+1,0}^r$ and $G_{0,0}^r$.

A Use Eqs. (B.7) and (B.8) to compute $G_{i,i}^{r,R(i)}$ and $G_{N+1,i}^{r,R(i)}$ successively for $i = N \dots 1$.

B Compute $G_{N+1,0}^r$ and $G_{0,0}^r$ as

$$G_{0,0}^r = \left((g_L^r)^{-1} - H_{0,1} G_{1,1}^{r,(1)} H_{1,0} \right)^{-1} \quad (\text{B.11})$$

and

$$G_{N+1,0}^r = G_{N+1,1}^{r,(1)} H_{1,0} G_{0,0}^r, \quad (\text{B.12})$$

Applications: From $G_{N+1,0}^r$ and $G_{0,0}^r$ it is possible to calculate the transmission and reflection amplitudes t and r from left to right from the Fisher-Lee relation, Eq. (C.31) and (C.32).

Algorithm B.2. Compute $G_{0,0}^r$, $G_{N+1,0}^r$, $G_{0,N+1}^r$, $G_{N+1,N+1}^r$

A Use Eqs. (B.7), (B.8), and (B.9) and the recursion relation

$$G_{N+1,N+1}^{r,R(i-1)} = G_{N+1,N+1}^{r,R(i)} + G_{N+1,i}^{r,R(i)} H_{i,i-1} G_{i-1,N+1}^{r,R(i-1)} \quad (\text{B.13})$$

to compute $G_{i,i}^{r,R(i)}$, $G_{N+1,i}^{r,R(i)}$, $G_{i,N+1}^{r,R(i)}$ and $G_{N+1,N+1}^{r,R(i)}$ successively for $i = N \dots 1$.

¹ For the calculation of densities it is enough to compute $G_{i,i}^<$, whereas for current densities additionally $G_{i,i\pm 1}^<$ is needed.

B Compute $G_{0,0}^r$, $G_{N+1,0}^r$, $G_{0,N+1}^r$ and $G_{N+1,N+1}^r$ from Eqs. (B.11), (B.12) and the relations

$$G_{0,N+1}^r = G_{0,0}^r H_{0,1} G_{1,N+1}^{r,R(1)} \quad (\text{B.14})$$

and

$$G_{N+1,N+1}^r = G_{N+1,N+1}^{r,R(1)} + G_{N+1,1}^{r,R(1)} H_{1,0} G_{0,0}^r. \quad (\text{B.15})$$

Applications: These four Green's functions are needed in order to calculate transmission both from left to right and, vice versa, from right to left. This is usually not necessary for two-terminal systems, where the transmission probabilities in both directions are connected by symmetries [59]. However, the computation of all four Green's functions is necessary for systems with multiple leads, where the left and right lead used in the RGF calculation may be composed of several real leads (see Section 4.4.2).

Algorithm B.3. Calculation of $G_{N+1,0}^r$ and $G_{i,i}^r$ for $i = 0 \dots N + 1$.

- A Use Eqs. (B.7) and (B.8) to compute $G_{i,i}^{r,R(i)}$ and $G_{N+1,i}^{r,R(i)}$ successively for $i = N \dots 1$.
- B Compute $G_{N+1,0}^r$ and $G_{0,0}^r$ from Eqs. (B.11) and (B.12).
- C Compute $G_{i,i}^r$ successively for $i = 1 \dots N + 1$ from the relation

$$G_{i+1,i+1}^r = G_{i+1,i+1}^{r,R(i+1)} + G_{i+1,i+1}^{r,R(i+1)} H_{i+1,i} G_{i,i}^r H_{i,i+1} G_{i+1,i+1}^{r,R(i+1)}. \quad (\text{B.16})$$

Applications: Computing these Green's functions allows at the same time a calculation of transport properties (transmission and reflection amplitudes from $G_{N+1,0}^r$ and $G_{0,0}^r$) and the local density of states, Eq. (2.55) (from $G_{i,i}^r$).

Algorithm B.4. Compute $G_{i,0}^r$ and $G_{i,N+1}^r$ for $i = 1 \dots N + 1$.

- A Compute $G_{i,i}^{r,R(i)}$ and $G_{i,N+1}^{r,R(i)}$ from Eqs. (B.7) and (B.9), successively from $i = N \dots 1$.
- B Compute $G_{0,0}^r$ and $G_{0,N+1}^r$ from Eqs. (B.11) and (B.14).
- C Compute $G_{i,0}^r$ and $G_{i,N+1}^r$ successively for $i = 1 \dots N + 1$ from the relations

$$G_{i+1,0}^r = G_{i+1,i+1}^{r,R(i+1)} H_{i+1,i} G_{i,0}^r \quad (\text{B.17})$$

and

$$G_{i+1,N+1}^r = G_{i+1,N+1}^{r,R(i+1)} + G_{i+1,i+1}^{r,R(i+1)} H_{i+1,i} G_{i,N+1}^r. \quad (\text{B.18})$$

Applications: $G_{i,0}^r$ and $G_{i,N+1}^r$ can be used for calculating transport (since they include $G_{0,0}^r$ and $G_{0,N+1}^r$) and for evaluating the lesser Green's function $G^<$ from Eq. (B.10).

B.3 Computational complexity and implementation

The recursive Green's function algorithms B.1 – B.4 involve multiplications and inversions of $M_i \times M_i$ matrices, where M_i is the size of the matrix block $H_{i,i}$. Both multiplication and inversion scale cubically with the matrix size, i.e. as M_i^3 . Each of the algorithms performs a constant number of operations for block i , and hence the computational complexity of all these algorithms is $\propto \sum_{i=0}^{N+1} M_i^3$. Thus, all algorithms scale equally with the system size, although of course the absolute computational time depends on the number of operations per block.

In contrast to the computational complexity that is the same for all algorithms, the storage requirements of these algorithms is quite different. In the algorithms B.1 and B.2 we need the values of block i only to calculate block $i - 1$. As this is done successively, i.e. the algorithms pass through the structure once from right to left, they require $\propto \max(M_i)^2$ memory. In contrast, the algorithms B.3 and B.4 additionally require the value of block i in order to calculate the full Green's function at i . In other words, these algorithms first pass through the structure from right to left and then back again from left to right. As a consequence, we need to store the intermediate results of the first pass for the second pass. Thus, algorithms B.3 and B.4 require $\propto \sum_{i=0}^{N+1} M_i^2$ memory. If the lack of memory becomes an issue, it would be possible to reduce the storage requirements using a technique similar to the “unknitting” step in Ref. [146].

Finally, a few words about the implementation. Once the Hamiltonian is in block-tridiagonal form, the RGF algorithms only involve straightforward matrix algebra. In order to implement the RGF method, we can therefore employ highly optimized linear algebra libraries. Nowadays, the LAPACK and BLAS interfaces [107] have become the standard in terms of linear algebra routines. Using these interfaces allows to use the most optimized linear algebra routines available for a given machine. Since the computational time is almost exclusively spent in linear algebra calculations, the performance gain through the use of an optimized library can be huge [155]. On standard workstations we therefore employ the optimized ATLAS libraries [145].

Appendix C

Generalized Fisher-Lee relation

C.1 Landauer-Büttiker formalism

A very intuitive approach to the transport problem, pioneered by Landauer, connects the conductance of a device to a scattering problem [88, 89, 231]. Here we will formulate the scattering problem in the language of a tight binding model.

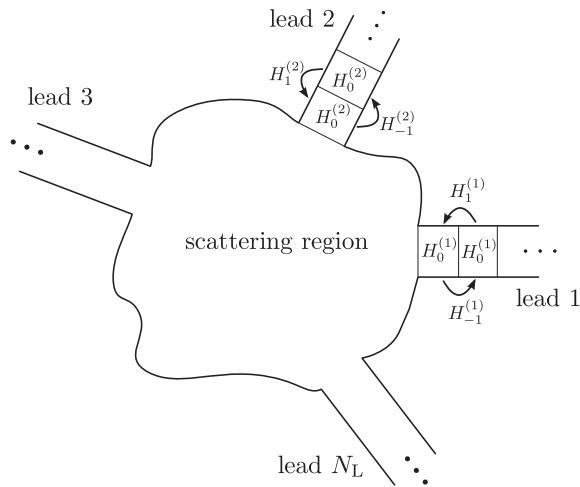


Figure C.1 – Schematic view of a scattering region connected to N_L leads. Note that the lead unit cells are labeled such that index 0 corresponds to the unit cell adjacent to the scattering region and increases when moving further into the lead.

Suppose a scattering region connected to N_L leads, as depicted in Fig. C.1. As in Chapter 3, we describe lead l as a semi-infinite repetition of unit cells, such that the Hamiltonian of the isolated lead is given by

$$H_{\text{semi-infinite}}^{(l)} = \sum_{j=0}^{\infty} H_0^{(l)} |j\rangle^{(l)} \langle j|^{(l)} + H_1^{(l)} |j\rangle^{(l)} \langle j+1|^{(l)} + H_{-1}^{(l)} |j\rangle^{(l)} \langle j-1|^{(l)}, \quad (\text{C.1})$$

where $H_0^{(l)}$ is the unit cell matrix and $H_{1,-1}^{(l)}$ are the hopping matrices. Note that we label the unit cells such that 0 denotes the unit cell adjacent to the scattering region, and the unit cell index increases as we move away from the scattering region (see Fig. C.1). The asymptotic eigenstates $\varphi_n^{(l)}$ for a given energy E and far away from the scattering region are given by the eigenstates of the corresponding infinite wire. They can be written in the form

$$\varphi_n^{(l)}(j) = \frac{1}{\sqrt{|v_n|}} \phi_n^{(l)} e^{ik_n j}, \quad (\text{C.2})$$

where $\phi_n^{(l)}$ is normalized to unity and obeys the equation

$$\left(E - H_0^{(l)} - H_1^{(l)} e^{ik_n} - H_{-1}^{(l)} e^{-ik_n} \right) \phi_n^{(l)} = 0, \quad (\text{C.3})$$

and v_n is the velocity of the eigenstate, Eq. 3.19. The asymptotic eigenstates $\varphi_n^{(l)}$ are called the *channels* of lead l . As in Chapter 3, we sort the channels according to their velocity v_n . For $v_n > 0$ we have a channel leaving the scattering region into the lead (*outgoing*, “<”), whereas for $v_n < 0$ the channel enters the scattering region from the lead (*incoming*, “>”). Note that we choose the notation “>” and “<” just as in Chapter 3.

A scattering-wave state $\Psi_n^{(l)}$ for a given energy E originates from an incoming channel $\varphi_{n,>}^{(l)}$ and is scattered into outgoing channels in the different leads attached to the scattering region. In the asymptotic region, where all evanescent waves have decayed, the scattering state can be written as

$$\Psi_n^{(l)}(j) = \begin{cases} \varphi_{n,>}^{(l)}(j) + \sum_m r_{ll,mn} \varphi_{m,<}^{(l)}(j) & \text{if } j \text{ in lead } l \\ \sum_{l'} t_{l'l,mn} \varphi_{m,<}^{(l')}(j) & \text{if } j \text{ in lead } l' \end{cases}. \quad (\text{C.4})$$

With the choice of wave function normalization for the channels in Eq. C.2, the transmission and reflection coefficients $t_{l'l,mn}$ and $r_{ll,mn}$ form a unitary scattering matrix [231]. The conductance coefficients for leads l' and l are then given by the Landauer-Büttiker formula (for further introductory reading, see Refs. [59, 60]):

$$G_{ll'} = \frac{e^2}{h} \sum_{n,m} |t_{l'l,mn}|^2. \quad (\text{C.5})$$

The problem of transport is thus reduced to calculating the scattering wave functions. However, very often it is more convenient to calculate the Green's function instead of the scattering wave functions itself. We are then left with the problem to obtain the transmission coefficients from the Green's function alone. Relations for the transmission or reflection coefficients and the Green's function of the system are called *Fisher-Lee* relations [90]. Here we will derive a generalized Fisher-Lee relation valid for any tight-binding Hamiltonian.

To this end we will closely follow Ref. [154], that derived Fisher-Lee relations for electrons in a magnetic field. The steps there carry over to arbitrary tight-binding Hamiltonians. Note that the problem of extracting the transmission coefficients from the Green's function has also been addressed in Ref. [82], utilizing the inverse matrix of eigenvectors ($U_{<}^{-1}$ in Chapter 3). The relations derived in this chapter avoid inverting the eigenvector matrix which can be problematic, as shown in Chapter 3. Furthermore, the expressions are more along the lines of the well-known Fisher-Lee relations for simple systems [59, 60, 90, 231], and will provide some additional physical insight.

The derivation of the generalized Fisher-Lee relation is complicated by the fact that the transverse wave functions of different channels need not be orthogonal any more, as discussed in Section 3.1. If the transverse wave functions were orthogonal, the transmission and reflection coefficients could be simply extracted from the scattering wave function by projection. To remedy the absence of a simple orthogonality relation, we derive a generalized orthogonality relation for the transverse wave functions ϕ_n in the next section. Then, we will construct the scattering wave function from the Green's function and derive expressions for the transmission and reflection coefficients. Finally, we emphasize the equivalence between the scattering formalism and the non-equilibrium Green's function formalism by showing how central results of both formalisms are connected.

C.2 Generalized orthogonality relations

We now proceed to derive a generalized orthogonality relation for the transverse wave functions of propagating and decaying asymptotic eigenstates \mathbf{u}_n .¹ Note that the expressions derived here are valid within a single lead l , and we will therefore drop the lead index “ (l) ” in this section in order to simplify notation.

By combining Eq. (3.14a) and its complex conjugate,

$$(E - H_0)\lambda_n \mathbf{u}_n = (H_1\lambda_n^2 + H_{-1}) \mathbf{u}_n \quad \text{and} \\ \mathbf{u}_n^\dagger (E - H_0)\lambda_n^* = \mathbf{u}_n^\dagger (H_{-1}(\lambda_n^*)^2 + H_1), \quad (\text{C.6})$$

we find that

$$\mathbf{u}_{n'}^\dagger (H_1\lambda_n^2 + H_{-1}) \mathbf{u}_n \lambda_n^* = \mathbf{u}_{n'}^\dagger (H_{-1}(\lambda_{n'}^*)^2 + H_1) \mathbf{u}_n \lambda_n \quad (\text{C.7})$$

¹ Remember that \mathbf{u}_n denotes all eigenstates, decaying and propagating, whereas ϕ_n denotes only propagating eigenstates.

and thus

$$\mathbf{u}_{n'}^\dagger (H_1 \lambda_n - H_{-1} \lambda_{n'}^*) \mathbf{u}_n (1 - \lambda_n \lambda_{n'}^*) = 0. \quad (\text{C.8})$$

This equation is trivially fulfilled if $\lambda_n \lambda_{n'}^* = 1$. This is the case for propagating modes with $\lambda_n = \lambda_{n'} = e^{ik_n}$. If $n = n'$, the expression in the first bracket is then up to a factor the velocity v_n of the propagating mode. For $n \neq n'$, i.e. for degenerate propagating modes, we have shown in Section 3.2.3, that the eigenmodes ϕ_n can be chosen such that they diagonalize the expression in the first bracket. In the case of decaying modes, the equation is trivially fulfilled, if $\lambda_n = 1/\lambda_{n'}^*$, for pairs of right- (“<”) and left-decaying (“>”) modes. In all other cases, the expression in the first bracket must be zero. In summary, we find

$$i \mathbf{u}_{n'}^\dagger (H_1 \lambda_n - H_{-1} \lambda_{n'}^*) \mathbf{u}_n = \begin{cases} \hbar v_n & \text{for propagating modes with } n = n' \\ \text{const.} & \text{for pairs of decaying modes with } \lambda_n = 1/\lambda_{n'}^* \\ 0 & \text{else} \end{cases}, \quad (\text{C.9})$$

where we have introduced a factor i to make the expression Hermitian. First, this proves that the propagating modes ϕ_n diagonalize the velocity operator, as stated in Section 3.2.3.² In addition, we are also provided with relations between propagating and decaying modes.

In order to utilize these relations, we introduce the operator

$$K(j) = \frac{i}{\hbar} (H_1 |j\rangle \langle j+1| - H_{-1} |j+1\rangle \langle j|). \quad (\text{C.10})$$

The full wave function of a decaying or propagating asymptotic state can be written in a compact form as $\xi_n(j) = c \mathbf{u}_n \lambda_n^j$, where c is an appropriate normalization constant ($c = 1/\sqrt{|v_n|}$ for propagating modes). Note that we reserve the notation $\varphi_n(j)$ for propagating modes. Then we find the generalized orthogonality relations:

$$\langle \xi_{n,<(>)} | K(j) | \xi_{n',<(>)} \rangle = \begin{cases} \delta_{nn'} & \text{for a propagating mode} \\ 0 & \text{else} \end{cases}, \quad (\text{C.11a})$$

and in particular for the propagating modes:

$$\langle \varphi_{n,<(>)} | K(j) | \varphi_{n',<(>)} \rangle = \delta_{nn'} \quad (\text{C.11b})$$

$$\langle \varphi_{n,>(<)} | K(j) | \xi_{n',<(>)} \rangle = 0, \quad (\text{C.11c})$$

regardless of j .

² Note furthermore, that if the transverse wave functions ϕ_n of the propagating modes would not diagonalize the velocity operator, the Landauer-Büttiker formula (C.5) could also contain terms of the form $t_{m_1 n_1} t_{m_2 n_2}^*$ with $m_1 \neq m_2$ and $n_1 \neq n_2$. This is an additional reason why in a numerical calculation degenerate propagating modes must be rotated so that the velocity operator is diagonal in this basis, as described in Section 3.2.3.

For the purposes of implementing these operator expressions in a computer program, it is useful to rewrite them in compact matrix form. For this, note that from the expressions for the surface Green's function in Section 3.2.4 we obtain:

$$g_R^r H_{-1} \mathbf{u}_{n,<} = \lambda_{n,<} \mathbf{u}_{n,<} \quad \text{and} \quad (\text{C.12a})$$

$$\mathbf{u}_{n,<}^\dagger H_1 g_R^a = \lambda_{n,<}^* \mathbf{u}_{n,<}^\dagger. \quad (\text{C.12b})$$

Expressions of the type $g_R^r H_{-1}$ essentially propagate a mode from one unit cell to the other and are thus commonly called *transfer matrices* [97].

Using the definition $\Gamma = i(\Sigma^r - \Sigma^a)$ with the lead self-energies $\Sigma^{r(a)} = H_1 g_R^{r(a)} H_{-1}$, we find

$$\begin{aligned} \mathbf{u}_{n',<} \Gamma \mathbf{u}_{n,<} &= i \mathbf{u}_{n',<} (H_1 g_R^r H_{-1} - H_1 g_R^a H_{-1}) \mathbf{u}_{n,<} \\ &= i \mathbf{u}_{n',<} (H_1 \lambda_{n,<} - H_{-1} \lambda_{n',<}^*) \mathbf{u}_{n,<} \\ &= \begin{cases} \hbar |v_n| \delta_{nn'} & \text{for propagating modes,} \\ 0 & \text{else.} \end{cases} \end{aligned} \quad (\text{C.13})$$

The matrix Γ thus plays the role of a velocity or current operator within the lead. In an analogous fashion, by utilizing the expressions for g_R^a , we can derive a similar expression for modes decaying to the right and propagating in the opposite direction.

The matrix Γ is a central object that appears in many expressions of the non-equilibrium Green's function formalism (see Section 2.5.2). In order to make connections between this formalism and the scattering theory, we collect some useful identities for Γ .

First, we rewrite the expression for the transfer matrix $g_R^r H_{-1}$ as

$$g_R^r H_{-1} = U_< \Lambda_< U_<^{-1} = \sum_n \mathbf{u}_{n,<} \lambda_{n,<} \tilde{\mathbf{u}}_{n,<}^\dagger, \quad (\text{C.14})$$

where $\tilde{\mathbf{u}}_{n,<}^\dagger$ is the n -th row of the inverse eigenvector matrix $U_<^{-1}$. Inserting this expression in the definition of the matrix Γ we obtain

$$\begin{aligned} \Gamma &= i (H_1 g_R^r H_{-1} - H_1 g_R^a H_{-1}) \\ &= i \left(H_1 U_< \Lambda_< U_<^{-1} - (U_<^\dagger)^{-1} \Lambda_<^\dagger U_<^\dagger H_{-1} \right) \\ &= i \sum_n \left(H_1 \mathbf{u}_{n,<} \lambda_{n,<} \tilde{\mathbf{u}}_{n,<}^\dagger - \tilde{\mathbf{u}}_{n,<} \lambda_{n,<}^* \mathbf{u}_{n,<}^\dagger H_{-1} \right), \end{aligned} \quad (\text{C.15})$$

where we utilized $(g_R^r H_{-1})^\dagger = H_1 g_R^a$. Now we insert unity at the appropriate places, $\mathbb{1} = U_< U_<^{-1} = \sum_m \mathbf{u}_{m,<} \tilde{\mathbf{u}}_{m,<}^\dagger$ and $\mathbb{1} = (U_<^\dagger)^{-1} U_<^\dagger = \sum_m \tilde{\mathbf{u}}_{m,<} \mathbf{u}_{m,<}^\dagger$. Then we find

$$\begin{aligned} \Gamma &= i \sum_n \left(\mathbb{1} H_1 \mathbf{u}_{n,<} \lambda_{n,<} \tilde{\mathbf{u}}_{n,<}^\dagger - \tilde{\mathbf{u}}_{n,<} \lambda_{n,<}^* \mathbf{u}_{n,<}^\dagger H_{-1} \mathbb{1} \right) \\ &= i \sum_{n,m} \left(\tilde{\mathbf{u}}_{m,<} \mathbf{u}_{m,<}^\dagger H_1 \mathbf{u}_{n,<} \lambda_{n,<} \tilde{\mathbf{u}}_{n,<}^\dagger - \tilde{\mathbf{u}}_{m,<} \lambda_{m,<}^* \mathbf{u}_{m,<}^\dagger H_{-1} \mathbf{u}_{n,<} \tilde{\mathbf{u}}_{n,<}^\dagger \right) \\ &= \sum_n \tilde{\phi}_{n,<} \hbar |v_{n,<}| \tilde{\phi}_{n,<}^\dagger, \end{aligned} \quad (\text{C.16})$$

where we used Eq. (C.9) and $|v_{n,<}| = v_{n,<}$. Furthermore, according to Eq. (C.9)

$$\tilde{\phi}_{n,<}^\dagger = \phi_{n,<}^\dagger \Gamma / (\hbar |v_{n,<}|) \quad (\text{C.17})$$

and we obtain the final expression

$$\Gamma = \sum_n \Gamma \phi_{n,<} \frac{1}{\hbar |v_{n,<}|} \phi_{n,<}^\dagger \Gamma. \quad (\text{C.18})$$

Only outgoing propagating modes $\phi_{n,<}$ contribute to the velocity/current operator Γ which comes at no surprise, since only the propagating modes carry current. From Eq. (C.17) and the eigendecomposition expression for the transfer matrix, we immediately obtain the useful relation

$$\phi_{n,<}^\dagger \Gamma g_R^r H_{-1} = \lambda_{n,<} \phi_{n,<}^\dagger \Gamma \quad (\text{C.19})$$

which is the complement to Eq. (C.12a).

In an analogous way, utilizing the eigenvector expression for the advanced surface Green's function, we can derive an expression for Γ only in terms of incoming propagating modes,

$$\Gamma = \sum_n \Gamma \phi_{n,>} \frac{1}{\hbar |v_{n,>}|} \phi_{n,>}^\dagger \Gamma. \quad (\text{C.20})$$

Note that $v_{n,>} < 0$ here. For the transfer matrices, we obtain

$$g_R^a H_{-1} \phi_{n,>} = \lambda_{n,>} \phi_{n,>} \quad \text{and} \quad (\text{C.21a})$$

$$H_1 g_R^r \Gamma \phi_{n,>} = \lambda_{n,>}^* \Gamma \phi_{n,>}. \quad (\text{C.21b})$$

C.3 Fisher-Lee relation

We now relate the scattering wave function $\Psi_n^{(l)}$ to the Green's function G^r . In order to keep the notation simple, we assume that the Hamiltonian of the full system can be written in a quasi-linear form,

$$H = \sum_{j=-\infty}^{\infty} H_{j,j} |j\rangle \langle j| + H_{j,j+1} |j\rangle \langle j+1| + H_{j,j-1} |j\rangle \langle j-1|. \quad (\text{C.22})$$

All blocks $j < 1$ are supposed to correspond to lead l , $|0\rangle = |0\rangle^{(l)}, |-1\rangle = |1\rangle^{(l)}, |-2\rangle = |2\rangle^{(l)}, \dots$, and all blocks $j > L$ are supposed to contain all other leads, $|L+1\rangle = \bigotimes_{l' \neq l} |0\rangle^{(l')}, |L+2\rangle = \bigotimes_{l' \neq l} |1\rangle^{(l')}, \dots$. The Hamiltonian H can always be written in such a form, and this form can be explicitly found by the algorithm presented in Chapter 4.

We now choose a finite region of integration \mathcal{A} containing the scattering region and a sizeable part of the leads, such that the boundary of \mathcal{A} is in the asymptotic region of the

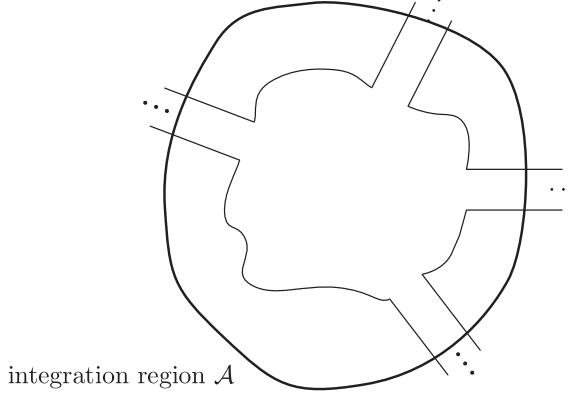


Figure C.2 – The integration region \mathcal{A} includes the scattering region and parts of the leads up to the asymptotic region.

leads, as schematically depicted in Fig. C.2. Using $G^r(E - H) = \mathbf{1}$ and $(E - H)\Psi_n^{(l)} = 0$ we can write

$$\Psi_n^{(l)}(\mathbf{x}) = \sum_{j \in \mathcal{A}} \langle \mathbf{x} | G^r(E - H) | j \rangle \langle j | \Psi_n^{(l)} \rangle - \langle \mathbf{x} | G^r | j \rangle \langle j | (E - H) | \Psi_n^{(l)} \rangle, \text{ for } \mathbf{x} \in \mathcal{A}. \quad (\text{C.23})$$

Using the quasi-linear form of H , Eq. (C.22), and writing accordingly $\sum_{j \in \mathcal{A}}$ as $\sum_{j=z_1}^{z_2}$, where $z_1 < 0$ and $z_2 > L + 1$ are blocks corresponding to the boundaries of \mathcal{A} , we find that

$$\begin{aligned} \Psi_n^{(l)}(\mathbf{x}) = & \langle \mathbf{x} | G^r | z_2 \rangle H_{z_2, z_2+1} \langle z_2 + 1 | \Psi_n^{(l)} \rangle - \\ & \langle \mathbf{x} | G^r | z_2 + 1 \rangle H_{z_2+1, z_2} \langle z_2 | \Psi_n^{(l)} \rangle + \\ & \langle \mathbf{x} | G^r | z_1 \rangle H_{z_1, z_1-1} \langle z_1 - 1 | \Psi_n^{(l)} \rangle - \\ & \langle \mathbf{x} | G^r | z_1 - 1 \rangle H_{z_1-1, z_1} \langle z_1 | \Psi_n^{(l)} \rangle. \end{aligned} \quad (\text{C.24})$$

The sum over all lattice points within the area \mathcal{A} has thus been reduced to a sum involving only contributions from the surface. Hence, this can be understood as a discrete version of Green's theorem. Using the definition from Eq. (C.10) we finally arrive at

$$\Psi_n^{(l)}(\mathbf{x}) = \sum_{l' \neq l} \frac{\hbar}{i} \langle \mathbf{x} | G^r K^{(l')}(z_2) | \Psi_n^{(l)} \rangle + \frac{\hbar}{i} \langle \mathbf{x} | G^r K^{(l)}(z_1) | \Psi_n^{(l)} \rangle. \quad (\text{C.25})$$

The expression for the scattering wave function $\Psi_n^{(l)}$ thus involves the operator $K^{(l)}$ for which we derived generalized orthogonality relations in the previous section. To simplify this expression further, we consider the behavior of the Green's function in the

asymptotic region of a lead, $\lim_{z \rightarrow \pm\infty} \langle \mathbf{x} | G^r | z \rangle$. Here, $\lim_{z \rightarrow \pm\infty}$ should be understood as taking z into the asymptotic region of lead l ($z \rightarrow -\infty$) and $l' \neq l$ ($z \rightarrow +\infty$). The Green's function $\lim_{z \rightarrow \pm\infty} \langle \mathbf{x} | G^r | z \rangle$ contains only incoming waves $\varphi_>$ (for a detailed discussion see Ref. [154]). Considering further that the scattering wave function $\Psi_n^{(l)}(z)$ contains incoming waves only in lead l , by virtue of the generalized orthogonality relations all surface terms except for the source term in lead l vanish:³

$$\begin{aligned} \Psi_n^{(l)}(\mathbf{x}) &= \lim_{z_1 \rightarrow -\infty} \langle \mathbf{x} | G^r | z_1 \rangle^{(l)} H_1^{(l)} \langle z_1 - 1 |^{(l)} | \Psi_n^{(l)} \rangle - \\ &\quad \langle \mathbf{x} | G^r | z_1 - 1 \rangle^{(l)} H_{-1}^{(l)} \langle z_1 |^{(l)} | \Psi_n^{(l)} \rangle \\ &= \lim_{z_1 \rightarrow -\infty} \left(G_{\mathbf{x}, z_1}^r H_1^{(l)} \frac{1}{\sqrt{|v_{n,>}|}} \phi_{n,>}^{(l)} e^{-ik_{n,>} (z_1 - 1)} - \right. \\ &\quad \left. G_{\mathbf{x}, z_1 + 1}^r H_{-1}^{(l)} \frac{1}{\sqrt{|v_{n,>}|}} \phi_{n,>}^{(l)} e^{-ik_{n,>} z_1} \right) \\ &= \lim_{z_1 \rightarrow -\infty} G_{\mathbf{x}, z_1}^r \left(H_1 g_R^{a,(l)} H_{-1} - H_1 g_R^{r,(l)} H_{-1} \right) \frac{1}{\sqrt{|v_{n,>}|}} \phi_{n,>}^{(l)} e^{-ik_{n,>} z_1}, \end{aligned} \quad (\text{C.26})$$

where we used Eqs. (C.21a) and (3.8). Using the definition of Γ_l we thus arrive finally at

$$\Psi_n^{(l)}(\mathbf{x}) = \lim_{z_1 \rightarrow -\infty} \frac{i}{\sqrt{|v_{n,>}|}} e^{-ik_{n,>} z_1} G_{\mathbf{x}, 0}^r \Gamma_l \phi_{n,>}^{(l)} \quad (\text{C.27})$$

valid for *any* point $\mathbf{x} \in \mathcal{A}$. This expression is thus a viable way to calculate the full scattering wave function in the scattering region from the Green's function G^r .

As usual [90, 154, 231], the relation between the scattering wave function and the Green's function involves values of G^r in the asymptotic region, i.e. far from the scattering region. For numerics, this is rather inconvenient, as it implies including large parts of the leads in the calculation, increasing computing time. It would therefore be desirable to obtain an expression involving only the Green's function within the first few layers of the lead. For this, note that from Eq. (3.8) we can write

$$G_{z_2, z_1}^r = G_{z_2, z'_1}^r \left(H_1 g_R^{r,(l)} \right)^{z'_1 - z_1} \quad (\text{C.28})$$

whenever z_2 is not further in lead l than z'_1 . In particular, we find restricting x to be contained in a block $j \geq 0$

$$\Psi_n^{(l)}(\mathbf{x}) = \lim_{z_1 \rightarrow -\infty} \frac{i}{\sqrt{|v_{n,>}|}} e^{-ik_{n,>} z_1} G_{\mathbf{x}, 0}^r \left(H_1 g_R^{r,(l)} \right)^{-z_1} \Gamma_l \phi_{n,>}^{(l)} \quad (\text{C.29})$$

Since according to Eq. (C.21b) $\Gamma_l \phi_{n,>}^{(l)}$ is an eigenvector of the transfer matrix $H_1 g_R^{r,(l)}$ with eigenvalue $\lambda_{n,>}^* = e^{-ik_{n,>}}$, we finally arrive at

$$\Psi_n^{(l)}(\mathbf{x}) = \frac{i}{\sqrt{|v_{n,>}|}} G_{\mathbf{x}, 0}^r \Gamma_l \phi_{n,>}^{(l)}. \quad (\text{C.30})$$

³ When reading the formulas remember that $|z_1\rangle = |-z_1\rangle^{(l)}$ in our notation. Therefore, additional “-”-signs appear with respect to the definition of the scattering wave-function $\Psi_n^{(l)}$ in Eq. (C.4)

This expression can be used directly for numerical computations.

In order to calculate the transmission and reflection coefficients we do not need the scattering wave function in the scattering region, but only in the respective leads. In the asymptotic region, the scattering wave function $\Psi_n^{(l)}$ contains only propagating modes, as indicated by Eq. (C.4). However, as argued above, for numerical purposes it is rather inconvenient to use the asymptotic region.

Closer to the scattering region, $\Psi_n^{(l)}$ also contains evanescent waves, decaying away in the direction of the lead. In the previous section we have shown that $\phi_{m,>}^{(l')}$ is orthogonal to both propagating and decaying modes with respect to the current operator $\Gamma_{l'}$. Therefore, we can use this generalized orthogonality relation to project out the desired scattering coefficients from the scattering wave function $\Psi_n^{(l)}$ at any point of the leads.

We first consider the case that \mathbf{x} is in lead $l' \neq l$. From $(\phi_{m,>}^{(l')})^\dagger \Gamma_{l'} \Psi_n^{(l)}$ we find

$$t_{l'l,mn} = \frac{i}{\hbar} \frac{1}{\sqrt{|v_{m,<}||v_{n,>}|}} (\phi_{m,<}^{(l')})^\dagger \Gamma_{l'} G_{L+1,0}^r \Gamma_l \phi_{n,>}^{(l)}. \quad (\text{C.31})$$

In the case of $l' = l$, we obtain the reflection coefficients. In this case it is necessary to keep in mind that incoming and outgoing waves need not be orthogonal with respect to Γ_l , and we obtain

$$r_{ll,mn} = \frac{1}{\hbar \sqrt{|v_{m,<}||v_{n,>}|}} \left(i (\phi_{m,<}^{(l)})^\dagger \Gamma_l G_{0,0}^r \Gamma_l \phi_{n,>}^{(l)} - (\phi_{m,<}^{(l)})^\dagger \Gamma_l \phi_{n,>}^{(l)} \right). \quad (\text{C.32})$$

These general expressions for the transmission and reflection coefficients are very similar to the Fisher-Lee relations derived for the particular case of electrons in a two-dimensional electron gas in textbooks [59, 60]. There, in- and outgoing modes have the same transverse wave function, $\phi_{n,<} = \phi_{n,>} = \phi_n$, and the transverse wave functions are eigenstates to the current operator, $\Gamma \phi_n = \hbar v_n \phi_n$. In general, this is not the case, and one has to resort to the formulas derived here. Examples where this is the case are electrons in a magnetic field, dealt with in Chapter 3 of this work, and zigzag graphene nanoribbons, Chapters 6 and 7.

C.4 Equivalence of scattering and non-equilibrium Green's function formalism

The generalized Fisher-Lee relation allows to relate central results from the non-equilibrium Green's function (NEGF) formalism with concepts from the Landauer-Büttiker formalism based on a scattering problem.

In the NEGF formalism, Eq. (2.119), we obtained the conductance between two leads l and l' in linear response as

$$G_{ll'} = \frac{e^2}{h} \text{Tr} (\Gamma_l G^r \Gamma_{l'} G^a). \quad (\text{C.33})$$

Expressing the current operators $\Gamma^{(l),(l')}$ in terms of in- and outgoing modes, Eqs. (C.18) and (C.20) we find

$$\begin{aligned}
G_{ll'} &= \frac{e^2}{h} \text{Tr} (\Gamma_l G^r \Gamma_{l'} G^a) \\
&= \frac{e^2}{h} \text{Tr} \left(\sum_n \Gamma_l \phi_{n,<}^{(l)} \frac{1}{\hbar|v_{n,<}|} (\phi_{n,<}^{(l)})^\dagger \Gamma_l G^r \sum_m \Gamma_{l'} \phi_{m,>}^{(l')} \frac{1}{\hbar|v_{m,>}|} (\phi_{m,>}^{(l')})^\dagger \Gamma_{l'} G^a \right) \\
&= \frac{e^2}{h} \sum_{m,n} \frac{1}{\hbar^2 |v_{n,<}| |v_{m,>}|} \left((\phi_{n,<}^{(l)})^\dagger \Gamma_l G^r \Gamma_{l'} \phi_{m,>}^{(l')} \right) \left((\phi_{m,>}^{(l')})^\dagger \Gamma_{l'} G^a \Gamma_l \phi_{n,<}^{(l)} \right) \\
&= \frac{e^2}{h} \sum_{m,n} |t_{ll',nm}|^2,
\end{aligned} \tag{C.34}$$

where we made use of the Fisher-Lee relation (C.31). Thus, the NEGF expression for the conductance in linear response is equivalent to the Landauer-Büttiker formula.

Another important relation from the NEGF formalism is the expression for the lesser Green's function from the Keldysh equation, Eq. (2.95),

$$G^< = G^r \Sigma^< G^a. \tag{C.35}$$

In the case of non-interacting electrons, or for Coulomb interaction in Hartree-Fock approximation, $\Sigma^<$ only contains contributions from the leads and can be written as

$$\Sigma^<(E) = \sum_l i \Gamma_l f(E, \mu_l), \tag{C.36}$$

where μ_l is the electrochemical potential in lead l . The contribution of a single lead l to $G^<$ is given by

$$\begin{aligned}
G^{<,(l)}(\mathbf{x}, \mathbf{x}') &= G^r i \Gamma_l f(E, \mu_l) G^a \\
&= f(E, \mu_l) \sum_n i G^r \Gamma_l \phi_{n,>}^{(l)} \frac{1}{\hbar|v_{n,>}|} (\phi_{n,>}^{(l)})^\dagger \Gamma_l G^a \\
&\propto f(E, \mu_l) \sum_n i \Psi_n^{(l)}(\mathbf{x}) (\Psi_n^{(l)}(\mathbf{x}'))^* \frac{1}{\hbar|v_{n,>}|}.
\end{aligned} \tag{C.37}$$

Note that we neglected possible constant terms in the last line, in order to avoid discussing the appropriate normalization of the scattering wave functions. The term $1/|v_{n,>}|$ is up to a constant the one-dimensional density of states of the incoming channel n in lead l [89], and we denote this density of states as $D_n^{1d,(l)}(E)$. The full lesser Green's function thus reads

$$G^<(\mathbf{x}, \mathbf{x}', E) \propto \sum_l f(E, \mu_l) \sum_n i \Psi_n^{(l)}(\mathbf{x}) (\Psi_n^{(l)}(\mathbf{x}'))^* D_n^{1d,(l)}(E). \tag{C.38}$$

Hence, the lesser Green's function, and thus in particular the electron density for a given energy E , is composed of all scattering states, weighted with the respective density of states of the incoming channel and the Fermi function of the respective lead.

Appendix D

Details of the derivation of the lead Green's function

D.1 $\lambda(E + i\eta)$ for propagating modes

In the derivation of the general expression for the surface Green's function of a lead it is necessary to determine which poles contribute to the contour integral in Eq. (3.12). That is, we need to determine whether the eigenvalues λ_n of the quadratic eigenvalue equation

$$((E + i\eta - H_0)\lambda_n - H_1\lambda_n^2 - H_{-1}) \mathbf{u}_n = 0 \quad (\text{D.1})$$

are located inside or outside the unit circle. For any eigenvalue with $|\lambda_n| \neq 1$ this can be done without considering the infinitesimal shift $i\eta$. However, in the case of a propagating mode $\mathbf{u}_n = \phi_n$, we find that the eigenvalue $\lambda_n = e^{ik}$ and thus lies directly on the unit circle, if we neglect the infinitesimal shift $i\eta$. Therefore, it is vital to include $i\eta$ in the calculation for this case.

Numerically, this could be done by explicitly including some small but final imaginary part of the energy E in the calculation [102]. However, this just introduces a parameter into the numerical calculation that might need manual adjustment in order to achieve convergence. Therefore we will show in the following, how we can determine $\lambda_n(E + i\eta)$ for the solution including the infinitesimal shift $i\eta$ from the properties of the solution $\lambda_n(E)$, $\phi_n(E)$ of the quadratic eigenvalue problem *without* the infinitesimal $i\eta$:

$$((E - H_0)\lambda_n(E) - H_1\lambda_n(E)^2 - H_{-1}) \phi_n(E) = 0. \quad (\text{D.2})$$

In the course of this derivation we will also gain some insight into the arguments used in the derivation of the lead Green's function in Refs. [59, 82, 86] and justify these arguments mathematically.

In order to determine the solution including the infinitesimal shift $i\eta$ we use the Taylor expansion

$$\lambda_n(E + i\eta) = \lambda_n(E) + \frac{d\lambda_n}{dE} \Big|_E i\eta + \mathcal{O}(\eta^2). \quad (\text{D.3})$$

Note that a similar approaches have been used in Refs. [94, 100, 103].

In order to calculate $\frac{d\lambda_n}{dE}$ we take the derivative of Eq. (D.2) with respect to the energy:

$$\begin{aligned} & \left(\lambda_n(E) + (E - H_0) \frac{d\lambda_n}{dE} - H_1 2\lambda(E) \frac{d\lambda_n}{dE} \right) \phi(E) + \\ & ((E - H_0)\lambda_n(E) - H_1\lambda_n(E)^2 - H_{-1}) \frac{d\phi_n}{dE} = 0. \end{aligned} \quad (\text{D.4})$$

Using the fact that, as shown in Section 3.2.2, ϕ_n^\dagger is also a right eigenvector of the quadratic eigenvalue problem Eq. (D.2), we can get rid of the second term of Eq. (D.4) and find

$$\phi_n^\dagger(E) \left(\lambda_n(E) + (E - H_0) \frac{d\lambda_n}{dE} - H_1 2\lambda(E) \frac{d\lambda_n}{dE} \right) \phi(E) = 0. \quad (\text{D.5})$$

Further exploiting the fact that the ϕ_n are normalized, $\phi_n^\dagger \phi_n = 1$, we arrive at the expression

$$\frac{d\lambda_n}{dE} \Big|_E = \frac{\lambda_n(E)}{2\phi_n^\dagger H_1 \phi_n \lambda_n(E) - \phi_n^\dagger (E - H_0) \phi_n}. \quad (\text{D.6})$$

To determine whether $\lambda_n(E + i\eta)$ lies inside or outside the unit circle, we need to calculate $|\lambda_n(E + i\eta)|$. Using the Taylor expansion (D.3) we find

$$\begin{aligned} |\lambda_n(E + i\eta)| &= |\lambda_n(E) + \frac{d\lambda_n}{dE} i\eta + \mathcal{O}(\eta^2)| \\ &= \left(\lambda_n(E) + \frac{d\lambda_n}{dE} i\eta + \mathcal{O}(\eta^2) \right) \left(\lambda_n^*(E) - \left(\frac{d\lambda_n}{dE} \right)^* i\eta + \mathcal{O}(\eta^2) \right) \\ &= |\lambda_n(E)|^2 + i\eta \left(\lambda_n^*(E) \frac{d\lambda_n}{dE} - \lambda_n(E) \left(\frac{d\lambda_n}{dE} \right)^* \right) + \mathcal{O}(\eta^2) \\ &= 1 - 2\eta \Im \left(\lambda_n^*(E) \frac{d\lambda_n}{dE} \right) + \mathcal{O}(\eta^2). \end{aligned} \quad (\text{D.7})$$

Using Eq. (D.6) we find

$$\begin{aligned} \lambda_n^*(E) \frac{d\lambda_n}{dE} &= \frac{\lambda_n^*(E) \lambda_n(E)}{2\phi_n^\dagger H_1 \phi_n \lambda_n(E) - \phi_n^\dagger (E - H_0) \phi_n} \\ &= \frac{1}{\underbrace{2\phi_n^\dagger H_1 \phi_n \lambda_n(E)}_{\textcircled{1}} - \underbrace{\phi_n^\dagger (E - H_0) \phi_n}_{\textcircled{2}}}. \end{aligned} \quad (\text{D.8})$$

Since for real energy E —and only for real energies we can have propagating modes—the matrix $E - H_0$ is hermitian, term ② in the equation above is real, and only term ① contributes to the imaginary part. Furthermore, note that the imaginary part of term ① is related to the velocity of the propagating mode according to Eq. (3.19). Combining this information we finally arrive at

$$\Im \left(\lambda_n^*(E) \frac{d\lambda_n}{dE} \right) = - \underbrace{\frac{\Im (2\phi_n^\dagger H_1 \phi_n \lambda_n(E))}{|...|}}_{>0} = \frac{\hbar v_n(E)}{|...|}, \quad (\text{D.9})$$

and thus

$$|\lambda_n(E + i\eta)| = 1 - 2\hbar \underbrace{\frac{\eta}{|...|}}_{>0} v_n(E) + \mathcal{O}(\eta^2). \quad (\text{D.10})$$

In summary we find that for a propagating mode $\phi_n(E)$ $\lambda_n(E + i\eta)$ lies

- inside the unit circle, if $v_n(E) > 0$, i.e. for right-moving modes,
- inside the unit circle, if $v_n(E) < 0$, i.e. for left-moving modes.

This criterion only fails for $v_n = 0$, i.e. directly at a band edge. However, $v_n = 0$ also implies a diverging density of states and thus impedes a numerical calculation of the Green's function anyways. In such a situation it is advisable to move away a little bit from the band edge.

Refs. [59, 82, 86] directly invoke the concept of left- and right-going modes in order to fulfill the boundary conditions of retarded and advanced Green's function, i.e. in- and outgoing waves. The calculations of this section justify these assumptions mathematically.

D.2 Derivation of Eq. (3.27)

We shall now show how to derive Eq. (3.27). In Eq. (3.26) we have

$$G_{j,j'}^r = \frac{1}{2\pi i} \oint dz U_< z^{j-j'} (h(z))^{-1} \quad \text{for } j' \leq j, \quad (\text{D.11})$$

where we defined $h(z) = \tilde{H}(E + i\eta, z)U_<$.

In order to solve the contour integral of Eq. (D.11) we need to consider each component of the inverse matrix $h(z)^{-1}$ separately. The components of the inverse matrix are given by

$$(h(z)^{-1})_{ij} = \frac{1}{\det(h(z))} \text{adj}(h(z)), \quad (\text{D.12})$$

where $\text{adj}(h(z))$ is the adjugate of the matrix $h(z)$. $(h(z)^{-1})_{ij}$ is thus a rational function and we need to determine the zeroes of denominator and numerator and their multiplicities in order to solve the contour integral.

First, we consider the denominator. We have $\det(h(z)) = \det(\tilde{H}(E+i\eta, z)) \det(U_<)$. $\det(U_<)$ is simply a constant. Therefore we find

$$\det(h(z)) \propto \prod_n (z - \lambda_{n,<}) \prod_n (z - \lambda_{n,>}), \quad (\text{D.13})$$

as $\det(\tilde{H}(E+i\eta, z)) = 0$ for $z = \lambda_{n,</>}$. Note that for a rank-deficient hopping matrix H_1 , some of the eigenvalues are infinite. The expression above contains contains finite eigenvalues, and $\det(h(z))$ is a polynomial of degree $N + \text{rank}(H_1)$.

Next, we examine the numerator further. The adjugate of the matrix $h(z)$ is given by $\text{adj}(h(z))_{ij} = (-1)^{i+j} M_{ji}$ where M_{ji} is given by the determinant of the matrix resulting from deleting row j and column i from $h(z)$. Let $S_{N \setminus i}$ denote the set of permutations of the numbers $\{1, \dots, i-1, i+1, \dots, N\}$. Using the Leibniz formula for determinants we can write

$$\text{adj}(h(z))_{ij} = (-1)^{i+j} \sum_{\sigma \in S_{N \setminus i}} \text{sgn}(\sigma) \prod_{\substack{n=1 \\ n \neq j}}^N h(z)_{n\sigma(n)}. \quad (\text{D.14})$$

The matrix $h(z)$ has columns that have definite zeroes at $\lambda_{n,<}$: We can write

$$h(z) = \tilde{H}(E+i\eta, z) U_< = \left(\tilde{H}(E+i\eta, z) \mathbf{u}_{1,<} \Big| \dots \Big| \tilde{H}(E+i\eta, z) \mathbf{u}_{N,<} \right). \quad (\text{D.15})$$

Since $\tilde{H}(E+i\eta, z = \lambda_{n,<}) \mathbf{u}_{n,<} = 0$, the entries of the matrix $h_{ij}(z) = (\tilde{H}(E+i\eta, z) \mathbf{u}_{j,<})_i$ have a zero of at least multiplicity one at $\lambda_{j,<}$. Thus we find that

$$\text{adj}(h(z))_{ij} = C(z) \prod_{\substack{n=1 \\ n \neq i}}^N (z - \lambda_{n,<}), \quad (\text{D.16})$$

where $C(z)$ is some polynomial.

Combining the expressions for the numerator and the denominator we find

$$(h(z)^{-1})_{ij} = \frac{1}{z - \lambda_{i,<}} D(z) \prod_n \frac{1}{z - \lambda_{n,>}}, \quad (\text{D.17})$$

where $D(z)$ is some polynomial. Only the eigenvalues inside the unit circle, $\lambda_{n,<}$, contribute to the contour integral (D.11). Thus we can ignore the poles at $\lambda_{n,>}$ and absorb them in the definition of $D(z)$. In order to solve the integral we need to distinguish two cases:

- $D(\lambda_{i,<}) \neq 0$. In this case, $(h(z)^{-1})_{ij}$ has a simple pole at $\lambda_{i,<}$ and we find

$$\begin{aligned} \frac{1}{2\pi i} \oint dz z^{j-j'} (h(z)^{-1})_{ij} &= \frac{1}{2\pi i} \oint dz \frac{z^{j-j'} D(z)}{z - \lambda_{i,<}} \\ &= (\lambda_{i,<})^{j-j'} D(\lambda_{i,<}) \\ &= (\lambda_{i,<})^{j-j'} \frac{1}{2\pi i} \oint dz (h(z)^{-1})_{ij}. \end{aligned} \quad (\text{D.18})$$

Note that this equation also holds for $\lambda_{i,<} = 0$, when $z^{j-j'} (h(z)^{-1})_{ij}$ is an analytic function within the unit circle.

- $D(\lambda_{i,<}) = 0$. In this case, $(h(z)^{-1})_{ij}$ is an analytic function within the unit circle and

$$\frac{1}{2\pi i} \oint dz z^{j-j'} (h(z)^{-1})_{ij} = \frac{1}{2\pi i} \oint dz (h(z)^{-1})_{ij} = 0. \quad (\text{D.19})$$

Hence, we can still write

$$\frac{1}{2\pi i} \oint dz z^{j-j'} (h(z)^{-1})_{ij} = (\lambda_{i,<})^{j-j'} \frac{1}{2\pi i} \oint dz (h(z)^{-1})_{ij}. \quad (\text{D.20})$$

Finally, from Eqs. (D.18) and (D.20) we find

$$G_{j,j'}^r = \frac{1}{2\pi i} \oint dz U_< z^{j-j'} (h(z))^{-1} = \frac{1}{2\pi i} \oint dz U_< \Lambda_{<}^{j-j'} (h(z))^{-1} \quad \text{for } j' \leq j, \quad (\text{D.21})$$

where $\Lambda_<$ is a diagonal matrix with entries $(\Lambda_<)_{nn} = \lambda_{n,<}$. This concludes the proof of Eq. (3.27).

It should be noted that during the derivation it was assumed that $j' \leq j$. If $j' > j$, $z^{j-j'} = 1/z^{j'-j}$ introduces additional poles in the contour integral. However, by substituting $y = 1/z$ we can perform an analogous calculation, but in this case only the eigenvalues outside the unit circle contribute.

D.3 Summary of the numerical algorithms

D.3.1 Eigendecomposition based algorithms

The results from Section 3.2 allow for the computation of the surface Green's function for arbitrary tight-binding models. Here, we summarize the numerical algorithms, distinguishing the two cases of invertible and non-invertible hopping matrix H_1 .

Algorithm D.1. Invertible hopping matrix H_1

A From H_0 , H_1 , and H_{-1} build up the matrix

$$A = \begin{pmatrix} 0 & 1 \\ -(H_1)^{-1} H_{-1} & (H_1)^{-1} (E - H_0) \end{pmatrix}. \quad (\text{D.22})$$

B Solve the eigenproblem

$$A \begin{pmatrix} \mathbf{u}_n \\ \lambda_n \mathbf{u}_n \end{pmatrix} = \lambda_n \begin{pmatrix} \mathbf{u}_n \\ \lambda_n \mathbf{u}_n \end{pmatrix}, \quad (\text{D.23})$$

and extract the $2N$ eigenvectors \mathbf{u}_n and eigenvalues λ_n of the quadratic eigenproblem.

- C Identify degenerate eigenvalues of the form $\lambda = e^{ik}$ and rotate the corresponding eigenvectors by diagonalizing the velocity operator.
- D Build up the matrices $U_<$ and $\Lambda_<$ from the right-going eigenstates.
- E Calculate the surface Green's function as

$$g_R^r = U_< \Lambda_< U_<^{-1} (H_{-1})^{-1}. \quad (\text{D.24})$$

Implementation: The eigenvalue problem can be solved numerically by the LAPACK routine ZGEEV. LAPACK also provides standard routines for matrix multiplication and inversion [107].

Algorithm D.2. Singular hopping matrix H_1

- A From H_0 , H_1 , and H_{-1} build up the matrices

$$A = \begin{pmatrix} 0 & 1 \\ -H_{-1} & E - H_0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & H_1 \end{pmatrix} \quad (\text{D.25})$$

- B Solve the generalized eigenproblem

$$A \begin{pmatrix} \mathbf{u}_n \\ \lambda_n \mathbf{u}_n \end{pmatrix} = \lambda_n B \begin{pmatrix} \mathbf{u}_n \\ \lambda_n \mathbf{u}_n \end{pmatrix}, \quad (\text{D.26})$$

and extract the $2N$ eigenvectors \mathbf{u}_n and eigenvalues λ_n of the quadratic eigenproblem.

- C Identify degenerate eigenvalues of the form $\lambda = e^{ik}$ and rotate the corresponding eigenvectors by diagonalizing the velocity operator.
- D Build up the matrices $U_<$ and $\Lambda_<$ from the right-going eigenstates.
- E Calculate the surface Green's function as

$$g_R^r = (E - H_0 - H_1 U_< \Lambda_< U_<^{-1})^{-1} \quad (\text{D.27})$$

Implementation: The generalized eigenproblem can be solved numerically by the LAPACK routine ZGGEV [107]. Note that the numerical algorithms in LAPACK solve the generalized eigenvalue problem in the form

$$\beta A \mathbf{u} = \alpha B \mathbf{u}, \quad (\text{D.28})$$

such that $\lambda = \alpha/\beta$. In this way, even $\lambda = \infty$ is well-defined as $\beta = 0$.

D.3.2 Schur decomposition based algorithms

The results from Section 3.3 allow for a more stable calculation of the lead Green's function in terms of the (generalized) Schur decomposition. We now summarize the computational steps, distinguishing again the cases of invertible and singular hopping matrices H_1 .

The algorithms are valid as long as $\lambda_{n,<}$ and $\lambda_{n,>}$ do not contain the same eigenvalues. Otherwise, i.e. in situations as depicted in Fig. 3.3(c), the energy E must be shifted slightly before starting the calculation.

Algorithm D.3. Invertible hopping matrix H_1

A From H_0 , H_1 , and H_{-1} build up the matrix

$$A = \begin{pmatrix} 0 & 1 \\ -(H_1)^{-1} H_{-1} & (H_1)^{-1} (E - H_0) \end{pmatrix}. \quad (\text{D.29})$$

- B Calculate the Schur decomposition $Q^\dagger A Q = T$ and extract the eigenvalues λ_n .
- C Calculate the eigenvectors ϕ_n of A for all propagating modes, i.e. $|\lambda_n| = 1$
- D For those propagating modes, identify degenerate eigenvalues of the form $\lambda_n = \lambda_{n'}$ and rotate the corresponding eigenvectors by diagonalizing the velocity operator.
- E Classify the eigenvalues into left- and right-going using the calculated eigenvectors ϕ_n for propagating modes.
- F Reorder the Schur decomposition such that the eigenvalues $\lambda_{n,<}$ appear on the leading diagonal positions of the triangular matrix T , and update Q accordingly.
- G Calculate the surface Green's function as

$$g_R^r = Q_{21} Q_{11}^{-1} (H_{-1})^{-1}. \quad (\text{D.30})$$

Implementation: This algorithm can be fully implemented with routines from LAPACK [107]. The Schur decomposition is calculated using ZGEHRD, ZUNGHR and ZHSEQR. The eigenvectors for propagating modes are computed using inverse iteration [112] with ZHSEIN. Finally, the routine ZTRSEN is applied to reorder the Schur decomposition.

Algorithm D.4. Singular hopping matrix H_1

A From H_0 , H_1 , and H_{-1} build up the matrices

$$A = \begin{pmatrix} 0 & 1 \\ -H_{-1} & E - H_0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & H_1 \end{pmatrix} \quad (\text{D.31})$$

- B Calculate the generalized Schur decomposition $Q^\dagger A Z = T$ and $Q^\dagger B Z = S$, and extract the eigenvalues $\lambda_n = t_{nn}/s_{nn}$. Note that only the transformation Z needs to be calculated explicitly.
- C Calculate the eigenvectors ϕ_n for all propagating modes, i.e. $|\lambda_n| = 1$
- D For those propagating modes, identify degenerate eigenvalues of the form $\lambda_n = \lambda_{n'}$ and rotate the corresponding eigenvectors by diagonalizing the velocity operator.
- E Classify the eigenvalues into left- and right-going using the velocities calculated from the eigenvectors ϕ_n for propagating modes.

- F Reorder the generalized Schur decomposition such that the eigenvalues $\lambda_{n,<}$ appear on the leading diagonal positions of the triangular matrices T and S , and update Z accordingly.
- G Calculate the surface Green's function as

$$g_R^r = (E - H_0 - H_1 Z_{21} Z_{11}^{-1})^{-1} \quad (\text{D.32})$$

Implementation: Again, LAPACK provides all routines necessary for the computation. The generalized Schur decomposition is computed using `ZGGEV` and from that the eigenvectors for propagating modes using `ZTGEVC`. The routine `ZTGSEN` reorders the generalized Schur decomposition.

Appendix E

The Fiduccia-Mattheyses algorithm

E.1 Graphs and hypergraphs

The Fiduccia-Mattheyses algorithm was originally developed for hypergraph partitioning [132]. A hypergraph \mathcal{H} is an ordered pair $\mathcal{H} = (\mathcal{V}, \mathcal{N})$, where \mathcal{V} is a set of vertices, and \mathcal{N} a set of *nets* (also called *hyperedges*) between them. A net $n_i \in \mathcal{N}$ is a set of vertices, i.e. $n_i \subset \mathcal{V}$. An undirected graph is a special realization of a hypergraph, where every net contains exactly two vertices. Thus, any algorithm for hypergraph partitioning can also be applied to an undirected graph.

During the FM bisection, we have to consider the graph structure arising from the Hamiltonian matrix in order to minimize the number of cut edges (min-cut), whereas for minimizing the number of surface vertices, i.e. the number of cut nets (min-net-cut), the hypergraph structure arising from all nets $\text{net}(v)$ as defined in Eq. (4.11), $\mathcal{N} = \{\text{net}(v) \mid v \in \mathcal{V}\}$, is essential. For min-net-cut-min-cut optimization, we have to consider both structures simultaneously. A schematic representation of a graph and the corresponding hypergraph structure is shown in Fig. E.1.

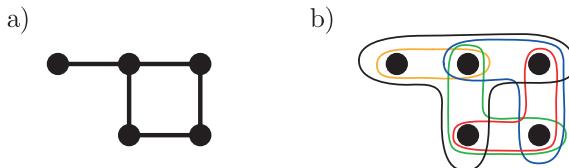


Figure E.1 – Schematic representation of (a) a simple graph and (b) the corresponding hypergraph structure imposed through all nets, $\mathcal{N} = \{\text{net}(v) \mid v \in \mathcal{V}\}$.

E.2 Fiduccia-Mattheyses bisection

The FM algorithm is based on the concept of *gain*. The gain of a vertex in an existing bisection is defined as the change in weight, i.e. the number of cut edges or nets, that occurs when this vertex is moved to the other part. This gain can also be negative, if such a move increases the number of cut edges or nets. The basic idea of the FM algorithm is to swap vertices with the highest gain between parts, while obeying some balance criterion. The fact that the highest gain can be negative, helps the FM algorithm to escape local minima. After moving, the respective vertex is locked in order to avoid an infinite loop, where a single vertex might be swapped back and forth repeatedly. The FM pass ends, when all (free) vertices have been moved, and the best bisection encountered during the pass is returned as result. Further passes can then successively improve on this bisection.

Appendix F

The method of finite differences

F.1 Basic ideas

The fundamental concept behind the method of finite differences go back to Newton and his contemporaries: Replace the derivative of a function by a difference quotient. With this replacement, a differential equation is reduced to a set of algebraic equations, that can easily be solved on a computer. In fact, the method of finite differences was already used for a numerical solution of the Schrödinger equation *before* the advent of electronic computers [77].

The approximations needed in the method of finite differences are most conveniently remembered from the Taylor expansion of a function f :

$$f(x + h) = f(x) + f'(x) h + \mathcal{O}(h^2) \quad (\text{F.1})$$

and

$$f(x - h) = f(x) - f'(x) h + \mathcal{O}(h^2) . \quad (\text{F.2})$$

Subtracting Eq. (F.2) from Eq. (F.1), we obtain the symmetric approximation for the derivative of a function,

$$f'(x) = \frac{f(x + h) - f(x - h)}{2h} + \mathcal{O}(h^2) , \quad (\text{F.3})$$

whereas adding Eqs. (F.1) and (F.2), we obtain an approximation for the function itself:

$$f(x) = \frac{1}{2} (f(x + h) + f(x - h)) + \mathcal{O}(h^2) . \quad (\text{F.4})$$

At first glance, Eq. (F.4) may not seem very useful. However, it can be used conveniently to bring the finite difference expressions into a symmetric form, yielding a

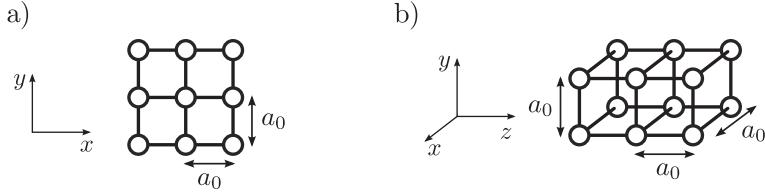


Figure F.1 – Grids used for the method of finite differences: (a) square grid with lattice constant a_0 in two dimensions. (b) cubic grid with lattice constant a_0 in three dimensions.

Hermitian Hamiltonian. Note that Eqs. (F.3) and (F.4) are valid up to second order in h . It is possible to use higher-order approximation formulas (see, for example, Ref. [232]), but we will restrict ourselves to the lowest order in this work.

In the method of finite differences, functions are not evaluated in the continuum, but only on a discrete set of grid points. The simplest grids in two and three dimensions are the uniform square and cubic grids with lattice constant a_0 , shown in Fig. F.1. Using Eqs. (F.3) and (F.4) with $h = a_0$ and $h = a_0/2$ when appropriate, it is possible to reduce a differential equation to a closed set of linear equations involving only function values on the grid points.

F.2 Example

In order to put the general ideas of the previous section into practice and to illustrate the typical steps in deriving a finite differences approximation to a Hamiltonian, we consider one example in great detail. Another detailed example can be found in the appendix of Ref. [155].

Consider the Hamiltonian H acting on the wave function $\varphi(x)$

$$\begin{aligned} H\varphi(x) &= \frac{1}{2}\left(\left(-i\hbar\partial_x + eA(x)\right)\alpha(x)\varphi(x) + \alpha(x)\left(-i\hbar\partial_x + eA(x)\right)\varphi(x)\right) \\ &= \frac{1}{2}\left(-i\hbar\left(\partial_x(\alpha(x)\varphi(x)) + \alpha(x)(\partial_x\varphi(x))\right) + 2e(A(x)\alpha(x)\varphi(x))\right). \end{aligned} \quad (\text{F.5})$$

Such terms arises from the Bychkov-Rashba Hamiltonian with a position-dependent Bychkov-Rashba parameter $\alpha(x)$ in a magnetic field with vector potential $A(x)$ (see Eq. (F.16)). Using (F.3), the first bracket in the last line of Eq. (F.5) can be rewritten as

$$\begin{aligned} \partial_x(\alpha(x)\varphi(x)) + \alpha(x)(\partial_x\varphi(x)) &\approx \\ \frac{1}{2a_0}\left((\alpha(x) + \alpha(x + a_0))\varphi(x + a_0) - (\alpha(x) + \alpha(x - a_0))\varphi(x - a_0)\right) & \end{aligned} \quad (\text{F.6})$$

The second bracket in the last line of Eq. (F.5) is conveniently recast into a more

symmetric form as

$$\begin{aligned} A(x)\alpha(x)\varphi(x) &= \frac{1}{4}\left((A(x)\alpha(x))\varphi(x) + A(x)(\alpha(x)\varphi(x))\right. \\ &\quad \left.+ \alpha(x)(A(x)\varphi(x)) + (A(x)\alpha(x)\varphi(x))\right) \\ &\approx \frac{1}{8}\left((\alpha(x) + \alpha(x + a_0))(A(x) + A(x + a_0))\varphi(x + a_0)\right. \\ &\quad \left.+ (\alpha(x) + \alpha(x - a_0))(A(x) + A(x - a_0))\varphi(x - a_0)\right), \end{aligned} \quad (\text{F.7})$$

where we used Eq. (F.4) on every one of the expressions in the inner brackets of the first line. Combining all terms, we finally arrive at

$$\begin{aligned} H\varphi(x) &\approx -\frac{i\hbar}{2a_0}\frac{1}{2}(\alpha(x) + \alpha(x + a_0))\varphi(x + a_0)\left(1 + \frac{iea_0}{\hbar}\frac{1}{2}(A(x) + A(x + a_0))\right) \\ &\quad + \frac{i\hbar}{2a_0}\frac{1}{2}(\alpha(x) + \alpha(x - a_0))\varphi(x - a_0)\left(1 - \frac{iea_0}{\hbar}\frac{1}{2}(A(x) + A(x + a_0))\right). \end{aligned} \quad (\text{F.8})$$

Introducing the notation $|x\rangle$ for a state at grid point x , we can write $H\varphi(x) = \langle x|H|\varphi\rangle$ and then read off the operator H from Eq. (F.8):

$$H \approx \sum_x -\frac{i\hbar}{2a_0}\frac{1}{2}(\alpha(x) + \alpha(x + a_0))e^{\frac{iea_0}{\hbar}\frac{1}{2}(A(x) + A(x + a_0))}|x\rangle\langle x + a_0| + \text{h.c.}, \quad (\text{F.9})$$

where h.c. denotes the Hermitian conjugate of the previous expression. In order to arrive at this expression, we have made use of the fact that all approximations made so far are valid up to second order in a_0 , and hence

$$1 + \frac{iea_0}{\hbar}\frac{1}{2}(A(x) + A(x + a_0)) \approx e^{\frac{iea_0}{\hbar}\frac{1}{2}(A(x) + A(x + a_0))}. \quad (\text{F.10})$$

The operator form of H in Eq. (F.9) has the form of a tight-binding operator. In this particular example, the Hamiltonian consists of a ‘‘hopping term’’ from point x to $x + a_0$, i.e. hopping between neighboring grid points. The magnetic field enters in the Hamiltonian in the form of a Peierls phase [233]

$$e^{\frac{iea_0}{\hbar}\frac{1}{2}(A(x) + A(x + a_0))} \approx e^{\frac{ie}{\hbar}\int_x^{x+a_0}ds \cdot \mathbf{A}(\mathbf{x})}. \quad (\text{F.11})$$

Because of this formal equivalence, the discretization procedure in the method of finite differences is said to yield a tight-binding representation of the Hamiltonian H .

F.3 Finite differences form of the Hamiltonian

After explicitly deriving the tight-binding representation of a particular example, we will now state the finite differences forms of several terms to the Hamiltonian encountered in this work. The expressions stated here are mostly valid in two ($\mathbf{x} = (x, y)$)

and in three dimensions ($\mathbf{x} = x, y, z$). For brevity, we will not explicitly write out coordinates that remain unchanged, i.e. do not contribute a hopping term. In addition, we will not explicitly include a magnetic field. In order to include the effect of a magnetic field, *every* hopping term must be multiplied by the Peierls phase (as in the previous example, see Eq. (F.11)). As a general guideline, if a Hamiltonian contains parameters that exhibit a discontinuity, the grid should be chosen such that the discontinuity lies halfway between two grid points [234].

Kinetic energy: The kinetic energy Hamiltonian with a position-dependent mass $m(\mathbf{x})$ [235] reads

$$H_{\text{kin}} = \frac{1}{2} \mathbf{p} \frac{1}{m(\mathbf{x})} \mathbf{p}, \quad (\text{F.12})$$

where $\mathbf{p} = -i\hbar\nabla$. The tight-binding representation is given as

$$\begin{aligned} H_{\text{kin}} = & \sum_{\mathbf{x}} \left(\frac{1}{m(x - \frac{a_0}{2})} + \frac{1}{m(x + \frac{a_0}{2})} \right) |x\rangle \langle x| \\ & - \left(\frac{1}{m(x + \frac{a_0}{2})} |x\rangle \langle x+a| + \text{h.c.} \right) \\ & + \text{analogous terms for the } y \text{ and } z\text{-direction.} \end{aligned} \quad (\text{F.13})$$

In the numerical calculation it is sometimes advantageous to substitute $m(x + \frac{a_0}{2}) \rightarrow \frac{1}{2}(m(x) + m(x + a_0))$ [234].

Potential energy: A potential enters the Hamiltonian as

$$H_{\text{pot}} = V(\mathbf{x}) \quad (\text{F.14})$$

and in tight-binding representation simply reads

$$H_{\text{pot}} = \sum_{\mathbf{x}} V(\mathbf{x}) |\mathbf{x}\rangle \langle \mathbf{x}|. \quad (\text{F.15})$$

Bychkov-Rashba spin-orbit coupling: The Bychkov-Rashba spin-orbit coupling Hamiltonian [21, 22] with a position-dependent Bychkov-Rashba parameter $\alpha(\mathbf{x})$ reads

$$H_{\text{BR}} = \frac{1}{2\hbar} \alpha(\mathbf{x}) (\sigma^x p_y - \sigma^y p_x) + \frac{1}{2\hbar} (\sigma^x p_y - \sigma^y p_x) \alpha(\mathbf{x}), \quad (\text{F.16})$$

where $\boldsymbol{\sigma}$ is the vector of Pauli matrices. Note that the Hamiltonian is symmetrized such that it is Hermitian (see Ref. [153, Chap. 3]). The tight-binding representation is then given as

$$\begin{aligned} H_{\text{BR}} = & \sum_{\mathbf{x}} \frac{-i}{2a_0} \sigma^x \frac{1}{2} (\alpha(y) + \alpha(y + a_0)) |y\rangle \langle y+a| + \text{h.c.} \\ & + \frac{i}{2a_0} \sigma^y \frac{1}{2} (\alpha(x) + \alpha(x + a_0)) |x\rangle \langle x+a| + \text{h.c.} \end{aligned} \quad (\text{F.17})$$

Dresselhaus spin-orbit coupling: The Dresselhaus spin-orbit coupling Hamiltonian [236] with a position-dependent Dresselhaus parameter $\gamma(z)$ reads

$$H_D = \frac{1}{\hbar} (\sigma^x p_x - \sigma^y p_y) \frac{\partial}{\partial z} \left(\gamma(z) \frac{\partial}{\partial z} \right) \quad (\text{F.18})$$

and in tight-binding representation

$$\begin{aligned} H_D = & \frac{-i}{2a_0^3} \sigma^x \sum_{\mathbf{x}} \gamma\left(z + \frac{a_0}{2}\right) |x, y, z\rangle \langle x + a_0, y, z + a_0| \\ & + \gamma\left(z - \frac{a_0}{2}\right) |x, y, z\rangle \langle x + a_0, y, z - a_0| \\ & - \left(\gamma\left(z - \frac{a_0}{2}\right) + \gamma\left(z + \frac{a_0}{2}\right) \right) |x, y, z\rangle \langle x + a_0, y, z| \\ & \frac{i}{2a_0^3} \sigma^y \sum_{\mathbf{x}} \gamma\left(z + \frac{a_0}{2}\right) |x, y, z\rangle \langle x, y + a_0, z + a_0| \\ & + \gamma\left(z - \frac{a_0}{2}\right) |x, y, z\rangle \langle x, y + a_0, z - a_0| \\ & - \left(\gamma\left(z - \frac{a_0}{2}\right) + \gamma\left(z + \frac{a_0}{2}\right) \right) |x, y, z\rangle \langle x, y + a_0, z| \\ & + \text{Hermitian conjugate of the whole expression.} \end{aligned} \quad (\text{F.19})$$

Note that in contrast to the previous expressions the tight-binding representation of the Dresselhaus parameter includes next-nearest neighbor hopping terms. Again, it may be useful to substitute $\gamma(z + \frac{a_0}{2}) \rightarrow \frac{1}{2}(\gamma(z) + \gamma(z + a_0))$.

F.4 Bloch's theorem and periodic boundary conditions

If a tight-binding Hamiltonian has a translational symmetry in some direction, Bloch's theorem can be used to reduce the size of the problem.

As example, consider a square grid, which is translationally invariant under a shift of $N_y a_0$, i.e. is composed of periodically repeated *supercells* with N_y lattice points in the y -direction (see Fig. F.2(a)). From Bloch's theorem, we can write

$$\varphi(x, y + N_y a_0) = e^{ik_y N_y a_0} \varphi(x, y), \quad (\text{F.20})$$

where $k_y \in [-\frac{\pi}{N_y a_0}, \frac{\pi}{N_y a_0}]$. In particular, this also holds for grid points at the boundary of the supercell. Let y_1 and y_2 denote the lower and upper row of grid points in the supercell. Then

$$\varphi(x, y_2 + a_0) = e^{ik_y N_y a_0} \varphi(x, y_1). \quad (\text{F.21})$$

Within the tight-binding Hamiltonian we can then replace hopping terms at the boundary as

$$|y_2\rangle \langle y_2 + a_0| \rightarrow e^{ik_y N_y a_0} |y_2\rangle \langle y_1| \quad (\text{F.22})$$

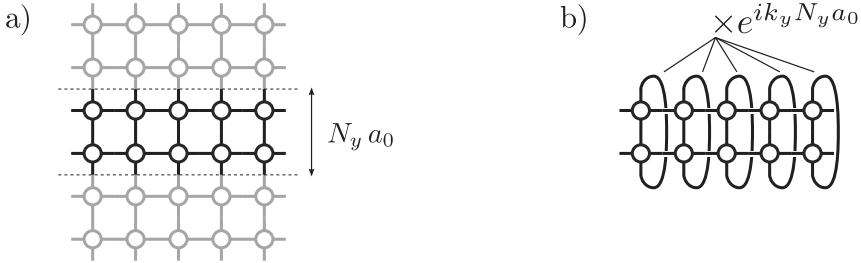


Figure F.2 – (a) A square grid consisting of a periodic repetition of supercells in the y -direction. (b) Effective structure after the application of Bloch’s theorem.

and

$$|y_1\rangle \langle y_1 - a_0| \rightarrow e^{-ik_y N_y a_0} |y_1\rangle \langle y_2|. \quad (\text{F.23})$$

Hence, the system size has an effective width of N_y grid points in the y -direction, however with additional hopping terms from the top to the bottom of the supercell and vice versa, with a phase given by the Bloch wave vector k_y .

Note that if we set $k_y = 0$, we obtain the conventional “periodic boundary conditions”. Neglecting all other Bloch vectors k_y is a good approximation, if $N_y a_0$ is large, i.e. k_y small. If this is not the case, the desired calculation must be performed for a set of k_y -values.

Appendix G

Tight-binding model for graphene

G.1 Lattice structure

Graphene is a two-dimensional realization of carbon, where the carbon atoms are arranged in a honeycomb network. The honeycomb network itself is not a Bravais lattice, but can be described as a triangular (or hexagonal) lattice with a basis of two atoms [93].

The triangular lattice is generated by the primitive basis vectors

$$\mathbf{a}_1 = \begin{pmatrix} a \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{pmatrix} \frac{1}{2}a \\ \frac{\sqrt{3}}{2}a \end{pmatrix}, \quad (\text{G.1})$$

as shown in Fig. G.1(a). The primitive unit cell spanned by these vectors is a rhombus with side length a . In the case of the honeycomb network, there are two basis atoms per unit cell at the positions

$$\mathbf{d}_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{d}_B = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{3}}a \end{pmatrix}. \quad (\text{G.2})$$

We will denote these basis atoms as atom A and B respectively. It should be emphasized that, although we label these basis atoms differently, within the infinitely extended honeycomb network, they are equivalent. The basis atoms form two triangular lattices (sublattice A and B, respectively), that are shifted with respect to each other to obtain the honeycomb network. In this network, each atom has three nearest neighbors of the opposite kind. The distance between neighboring carbon atoms is given as $a_{CC} = a/\sqrt{3}$.

In the case of a triangular lattice, the reciprocal is again triangular, rotated by 90° with respect to the original lattice, as shown in Fig. G.1(b). The reciprocal lattice is spanned by the primitive basis vectors [93]

$$\mathbf{b}_1 = \begin{pmatrix} \frac{2\pi}{a} \\ \frac{2\pi}{\sqrt{3}a} \end{pmatrix} \quad \text{and} \quad \mathbf{b}_2 = \begin{pmatrix} 0 \\ \frac{4\pi}{\sqrt{3}a} \end{pmatrix}. \quad (\text{G.3})$$

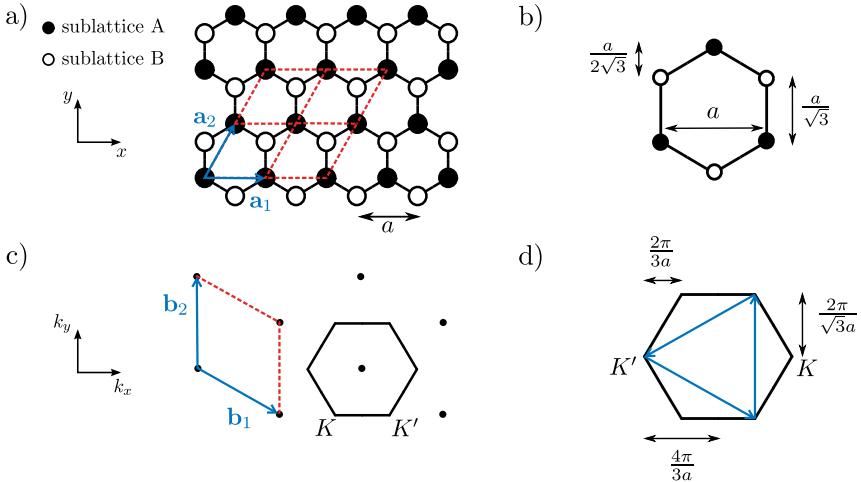


Figure G.1 – (a) Honeycomb network of carbon atoms. The underlying triangular Bravais lattice is spanned by the primitive basis vectors \mathbf{a}_1 and \mathbf{a}_2 , shown in blue. The primitive unit cell is a rhombus, as depicted by the red dashed lines. The two basis atoms are shown as black (sublattice A) and white (sublattice B) dots, respectively. (b) Geometrical details of a hexagon from the honeycomb network, formed by carbon atoms (c) The reciprocal lattice is spanned by the reciprocal basis vectors \mathbf{b}_1 and \mathbf{b}_2 , shown in blue. The first Brillouin zone has a hexagonal shape. (d) Geometrical details of the first Brillouin zone, showing the reciprocal lattice vectors connecting the three equivalent K' -corners.

The first Brillouin zone is constructed as the Wigner-Seitz cell of the reciprocal lattice [93] and has the shape of a hexagon. However, a wave vector in a Bravais lattice is only defined up to some reciprocal lattice vector. Thus, only two of the six corners in the Brillouin zone are non-equivalent, as always three corners can be connected with a reciprocal lattice vector. These two inequivalent corners are called K -points and are labeled K and K' respectively. For simplicity, we choose the wave vectors

$$\mathbf{K} = \begin{pmatrix} \frac{4\pi}{3a} \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{K}' = \begin{pmatrix} -\frac{4\pi}{3a} \\ 0 \end{pmatrix} \quad (\text{G.4})$$

to represent the two K -points.

It should be emphasized that the hexagonal shape of the Brillouin zone, and thus the existence of two K -points, is solely due to the underlying triangular Bravais lattice structure of the honeycomb network. The reciprocal lattice *only* depends on the Bravais lattice structure, not on the basis atoms per unit cell. In particular, the hexagonal shape of the Brillouin zone is *not* related to the hexagons formed by carbon atoms in the honeycomb network.

G.2 Electronic structure

G.2.1 Tight-binding model

We consider the dynamics of electrons in the graphene lattice assuming that the effect of interactions can be incorporated by a mean field (non-interacting electrons; for a justification see e.g. [237, Chap. 5]). In this case, the single-particle Hamiltonian of the system is given as

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V_{\text{eff}}(\mathbf{x}) + V_{\text{ext}}(\mathbf{x}), \quad (\text{G.5})$$

where $V_{\text{eff}}(\mathbf{x})$ incorporates both the potential arising from the ions within the graphene lattice and the effect of interaction in an approximate way, and V_{ext} is some external potential. Note that in the case of an infinite graphene sheet, $V_{\text{eff}}(\mathbf{x})$ has the full symmetry of the graphene lattice. We now intend to develop a tight-binding formulation for this problem, as was done first by Wallace in 1947 [238].

Carbon is the sixth element in the periodic table, and has 4 valence electrons in $2s^22p^2$ configuration. In graphene (and graphite, from which graphene originates), three of the valence electrons hybridize in a $2sp^2$ configuration. These electrons form bonds with the neighboring carbon atoms, leading to the honeycomb arrangement of carbon atoms. The fourth electron resides in a $2p_z$ -orbital perpendicular to the graphene plane. The electrons that form the bonds to the neighboring carbon atoms will not contribute to transport due to their bound nature, and we can therefore treat graphene as having one electron per carbon atom in the $2p_z$ state.

Following the arguments above, we describe graphene in tight-binding approximation with a single orbital per basis atom. The orbital at an atom situated at $\mathbf{R}_i + \mathbf{d}_\alpha$ is given as

$$\chi_{i,\alpha,s}(\mathbf{x}) = \chi_{2p_z,s}(\mathbf{x} - \mathbf{R}_i - \mathbf{d}_\alpha) \quad (\text{G.6})$$

where i denotes the (Bravais) lattice vector $\mathbf{R}_i = n\mathbf{a}_1 + m\mathbf{a}_2$ ($n, m \in \mathbb{Z}$), α the sublattice ($\alpha = \text{A}, \text{B}$) and s the spin. $\chi_{2p_z,s}(\mathbf{x} - \mathbf{R}_i - \mathbf{d}_\alpha)$ is the wave function of a $2p_z$ -orbital centered at atom α in the unit cell at the lattice point \mathbf{R}_i , with spin s . The quantum mechanical state corresponding to the wave function $\chi_{i,\alpha,s}(\mathbf{x})$ will be denoted as $|i, \alpha, s\rangle$. As discussed in Section 2.5.1, we assume that the states $|i, \alpha, s\rangle$ form an orthonormal set, thus neglecting overlap between wave functions located at different atoms. Within this basis, the tight-binding Hamiltonian reads

$$H = \sum_{i,\alpha,j,\beta,s,s'} \langle i, \alpha, s | H | j, \beta, s' \rangle |i, \alpha, s\rangle \langle j, \beta, s'|. \quad (\text{G.7})$$

In the remainder of this chapter we will only consider systems where spin is a good quantum number, and suppress the spin index s in the notation. In this case, spin only enters in a twofold occupation of energy levels. When we consider spin-dependent transport properties of graphene nanoribbons in Chapter 7, we will explicitly discuss how spin-dependent terms enter the Hamiltonian.

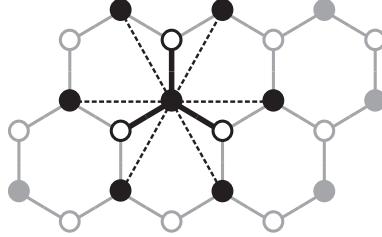


Figure G.2 – The nearest neighbors of an A atom are three B atoms (connections shown as solid black lines), the next-nearest neighbors are six A atoms (connections shown as dashed black lines).

In the spin-less case, the Hamiltonian reads

$$H = \sum_{i,\alpha,j,\beta} t_{i\alpha j\beta} |i, \alpha\rangle \langle j, \beta| , \quad (\text{G.8})$$

where $t_{i\alpha j\beta} = \langle i\alpha | H | j\beta \rangle$. The solution has the form

$$|\varphi\rangle = \sum_{i,\alpha} \varphi_\alpha(i) |i, \alpha\rangle , \quad (\text{G.9})$$

where $\varphi_\alpha(i)$ is the probability amplitude of occupying the orbital state $|i, \alpha\rangle$. In real-space, the wave function reads

$$\varphi(\mathbf{x}) = \sum_{i,\alpha} \varphi_\alpha(i) \chi_{2p_z}(\mathbf{x} - \mathbf{R}_i - \mathbf{d}_\alpha) . \quad (\text{G.10})$$

Thus, due to the localized nature of the atomic orbitals, $|\varphi_\alpha(i)|^2$ can be roughly interpreted as the probability of being close to the atomic site $\mathbf{R}_i + \mathbf{d}_\alpha$.

The fact that the atomic orbitals are localized is also the reason why only a few of the matrix elements $t_{i\alpha j\beta}$ are nonzero. Here, we will only consider matrix elements involving orbitals at the same site, and orbitals that are nearest or next-nearest neighbors.

The matrix elements involving only orbitals at the same atomic site can be written as $t_{i\alpha i\alpha} = \epsilon_0 + \epsilon_{i\alpha}$, where ϵ_0 arises from the lattice-periodic part of the Hamiltonian, $-\frac{\hbar^2}{2m} \nabla^2 + V_{\text{eff}}$, and $\epsilon_{i\alpha}$ from V_{ext} . Due to the lattice periodicity, ϵ_0 is the same for every atomic site and can thus be set to zero. For the external potential, it is useful to recast $\epsilon_{i\alpha}$ into a different form:

$$\epsilon_{i\alpha} = \begin{cases} V_i + M_i & \text{for } \alpha = \text{A}, \\ V_i - M_i & \text{for } \alpha = \text{B}. \end{cases} \quad (\text{G.11})$$

Here, V_i has the meaning of a potential that has the *same* value for *both* atoms in the unit cell, and M_i is a *staggered* potential, i.e. has *opposite* sign for sublattice A and

B. Such a staggered potential can thus arise from a potential varying rapidly on the inter-atomic scale. In addition, it is expected to arise naturally by placing graphene on a suitable substrate (breaking the symmetry between A and B atoms) [31], and due to antiferromagnetic ordering in zigzag graphene nanoribbons, as discussed in Chapter 7.

We now turn to the matrix elements involving orbitals at different sites, i.e. hopping matrix elements. As seen from Fig. G.2, every atom of a given type is surrounded by three nearest atoms of the opposite kind. The distance vectors from an A atom to the nearest B atoms are given by

$$\boldsymbol{\delta}_1 = \begin{pmatrix} 0 \\ \frac{a}{\sqrt{3}} \end{pmatrix}, \quad \boldsymbol{\delta}_2 = \begin{pmatrix} \frac{a}{2} \\ -\frac{a}{2\sqrt{3}} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\delta}_3 = \begin{pmatrix} -\frac{a}{2} \\ -\frac{a}{2\sqrt{3}} \end{pmatrix}, \quad (\text{G.12})$$

the distance vectors from an B atom to the closest A atoms as $-\boldsymbol{\delta}_1$, $-\boldsymbol{\delta}_2$, and $-\boldsymbol{\delta}_3$. The next-nearest neighbors of an atom of a certain kind are given by six atoms of the same kind, with distance vectors

$$\boldsymbol{\delta}'_1 = \mathbf{a}_1, \boldsymbol{\delta}'_2 = \mathbf{a}_2, \boldsymbol{\delta}'_3 = \mathbf{a}_2 - \mathbf{a}_1, \boldsymbol{\delta}'_4 = -\mathbf{a}_1, \boldsymbol{\delta}'_5 = -\mathbf{a}_2, \text{ and } \boldsymbol{\delta}'_6 = \mathbf{a}_1 - \mathbf{a}_2. \quad (\text{G.13})$$

Since the $2p_z$ -orbital is symmetric with respect to rotations in the plane of the graphene sheet, we can parameterize all matrix elements involving nearest and next-nearest neighbors with a single parameter t and t' , respectively.

In summary, we consider a tight-binding model for graphene in *next-nearest neighbor approximation* with matrix elements

$$t_{i\alpha j\beta} = \begin{cases} V_i + M_i & \text{if } i = j, \alpha = \beta = \text{A} \\ V_i - M_i & \text{if } i = j, \alpha = \beta = \text{B} \\ -t & \text{if } i, \alpha \text{ and } j, \beta \text{ are nearest neighbors,} \\ -t' & \text{if } i, \alpha \text{ and } j, \beta \text{ are next-nearest neighbors.} \end{cases} \quad (\text{G.14})$$

In the spirit of the Slater-Koster approach to tight-binding [239], we do not attempt to calculate the hopping matrix elements from first principles, but as parameters that should be adjusted to fit *ab initio* calculations. The numerical values of the hopping matrix elements employed in this work are $t = 2.7 \text{ eV}$ and $t' = 0.1t$ [183, 184]. The effects of a magnetic field can be included through the Peierls phase [233], by replacing

$$t_{i\alpha j\beta} \rightarrow t_{i\alpha j\beta} \times \exp \left(\frac{ie}{\hbar} \int_{\mathbf{R}_j + \mathbf{d}_\beta}^{\mathbf{R}_i + \mathbf{d}_\alpha} d\mathbf{s} \cdot \mathbf{A}(\mathbf{x}) \right), \quad (\text{G.15})$$

where $\mathbf{A}(\mathbf{x})$ is the vector potential.

Before we proceed to calculate the band structure of this tight-binding model, let us revisit a few of the approximations made in its derivation: First, we always assumed that $V_{\text{eff}}(\mathbf{x})$ has the full symmetry of the graphene lattice. This is not the case for a finite system, where especially atoms close to the boundary would feel a very different

effective lattice potential. In principle, this could result in quite different tight-binding parameters close to the boundary. However, we will neglect these effects when considering finite systems. This common approximation can be justified *a posteriori* by the observation that the results agree favorably with other approaches, such as DFT, that include the above mentioned effects of the boundaries (For an example, see the discussion of zigzag graphene nanoribbons in Chapter 7).

Second, the model as discussed above considers only up to next-nearest neighbor hopping. Reich *et al.* additionally include third-nearest neighbor hopping [190] on the grounds that the third-nearest neighbor distance is only slightly larger than the next-nearest neighbor distance. However, third-nearest neighbor hopping introduces only some corrections to the nearest-neighbor hopping, whereas next-nearest neighbor hopping introduces terms in the Hamiltonian that have different symmetry, as discussed below. These terms of different symmetry play a crucial role in the discussion of transport in graphene nanoribbons in Chapter 6. Therefore, the inclusion of next-nearest neighbors is essential, whereas third-nearest neighbor hopping only gives small corrections.

Finally, the tight-binding model is based on a single $2p_z$ -orbital per atomic site. In principle, it is possible to include all valence electrons on equal footing [240, 241]. However, in the energy range relevant for transport, the band structures obtained from these extended tight-binding models are practically indistinguishable from the results of the much simpler single-orbital tight-binding approach employed here.

G.2.2 Band structure in tight-binding approximation

After introducing the tight-binding model for graphene, we now calculate the band structure for a homogeneous infinite graphene sheet, with $M_i = M$, i.e. a constant staggered potential, and $V_i = V_0$. Then, the system is invariant under lattice translations,¹ and Bloch's theorem holds. For this case, we can apply standard techniques to calculate the band structure [93, Chap. 10].

To this end, we define the *tight-binding sums*²

$$\Phi_\alpha(\mathbf{k}, \mathbf{x}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_i} e^{i\mathbf{k}(\mathbf{R}_i + \mathbf{d}_\alpha)} \chi_{i,\alpha}(\mathbf{x}), \quad (\text{G.16})$$

¹ At first glance, the staggered, and thus spatially dependent potential M seems to break translational invariance. However, the staggered potential changes sign only on A and B atoms within a unit cell, and exhibits the *triangular* lattice periodicity.

² Note that there are different conventions in the literature on how to define a tight-binding sum for a lattice with more than one basis atom. For example, the definition $\Phi_\alpha(\mathbf{k}, \mathbf{x}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_i} e^{i\mathbf{k}\mathbf{R}_i} \chi_{i,\alpha}(\mathbf{x})$ is also used commonly [93, 237]. However, these different definitions are equivalent, as the corresponding wave functions only differ by a (\mathbf{k} -dependent) phase. Here we follow the conventions set by the original calculation of Wallace [238] and the work of Haldane [242].

where $\alpha = A, B$ and N is the number of unit cells. The tight-binding sums form an orthonormal set, i.e.

$$\langle \Phi_\alpha(\mathbf{k}) | \Phi_\beta(\mathbf{k}') \rangle = \delta_{\alpha,\beta} \delta_{\mathbf{k},\mathbf{k}'} , \quad (\text{G.17})$$

and obey Bloch's theorem. From the tight-binding sums, we make the ansatz

$$\varphi(\mathbf{k}, \mathbf{x}) = \varphi_A(\mathbf{k}) \Phi_A(\mathbf{k}, \mathbf{x}) + \varphi_B(\mathbf{k}) \Phi_B(\mathbf{k}, \mathbf{x}) \quad (\text{G.18})$$

for the eigenfunctions with Bloch vector \mathbf{k} . Using the orthonormality of the tight-binding sums, the Schrödinger equation $H\varphi(\mathbf{k}, \mathbf{x}) = E\varphi(\mathbf{k}, \mathbf{x})$ turns into an eigenproblem involving only the coefficients $\varphi_{A,B}(\mathbf{k})$:

$$H(\mathbf{k}) \begin{pmatrix} \varphi_A(\mathbf{k}) \\ \varphi_B(\mathbf{k}) \end{pmatrix} = \begin{pmatrix} H_{AA}(\mathbf{k}) & H_{AB}(\mathbf{k}) \\ H_{BA}(\mathbf{k}) & H_{BB}(\mathbf{k}) \end{pmatrix} \begin{pmatrix} \varphi_A(\mathbf{k}) \\ \varphi_B(\mathbf{k}) \end{pmatrix} = E(\mathbf{k}) \begin{pmatrix} \varphi_A(\mathbf{k}) \\ \varphi_B(\mathbf{k}) \end{pmatrix} . \quad (\text{G.19})$$

Here, $H_{\alpha\beta} = \langle \Phi_\alpha(\mathbf{k}) | H | \Phi_\beta(\mathbf{k}) \rangle$. In order to calculate the band structure $E(\mathbf{k})$, we now evaluate these matrix elements using the tight-binding model in next-nearest neighbor approximation of the preceding section. We obtain

$$\begin{aligned} \langle \Phi_A(\mathbf{k}) | H | \Phi_A(\mathbf{k}) \rangle &= \frac{1}{N} \sum_{\mathbf{R}_i, \mathbf{R}_j} e^{i\mathbf{k}(\mathbf{R}_j - \mathbf{R}_i)} \langle i, A | H | j, A \rangle \\ &= \frac{1}{N} \sum_{\mathbf{R}_i, \mathbf{R}_j} e^{i\mathbf{k}(\mathbf{R}_j - \mathbf{R}_i)} \left((V_0 + M) \delta_{\mathbf{R}_i, \mathbf{R}_j} - t' \sum_{l=1}^6 \delta_{\mathbf{R}_j - \mathbf{R}_i, \delta'_l} \right) \\ &= V_0 + M - t' \sum_{l=1}^6 e^{i\mathbf{k}\delta'_l} = V_0 + M - 2t' \sum_{l=1}^3 \cos(\mathbf{k}\delta'_l) , \end{aligned} \quad (\text{G.20})$$

where we used $\delta_4 = -\delta_1$, etc. in the last line. In the same fashion we find the matrix element involving only the B sublattice as

$$\langle \Phi_B(\mathbf{k}) | H | \Phi_B(\mathbf{k}) \rangle = V_0 - M - 2t' \sum_{l=1}^3 \cos(\mathbf{k}\delta'_l) . \quad (\text{G.21})$$

The off-diagonal matrix elements evaluate to

$$\begin{aligned} \langle \Phi_A(\mathbf{k}) | H | \Phi_B(\mathbf{k}) \rangle &= \frac{1}{N} \sum_{\mathbf{R}_i, \mathbf{R}_j} e^{i\mathbf{k}(\mathbf{R}_j + \mathbf{d}_B - \mathbf{R}_i)} \langle i, A | H | j, B \rangle \\ &= -t \sum_{l=1}^3 e^{i\mathbf{k}\delta_l} = -t \sum_{l=1}^3 \cos(\mathbf{k}\delta_l) + i \sin(\mathbf{k}\delta_l) , \end{aligned} \quad (\text{G.22})$$

and $H_{BA}(\mathbf{k}) = (H_{AB}(\mathbf{k}))^*$.

The tight-binding Hamiltonian in \mathbf{k} -space can thus be written in the form [242]

$$H(\mathbf{k}) = \left(V_0 - 2t' \sum_{l=1}^3 \cos(\mathbf{k}\delta'_l) \right) \mathbb{1} - \left(t \sum_{l=1}^3 \cos(\mathbf{k}\delta_l) \right) \sigma^x + \left(t \sum_{l=1}^3 \sin(\mathbf{k}\delta_l) \right) \sigma^y + M \sigma^z , \quad (\text{G.23})$$

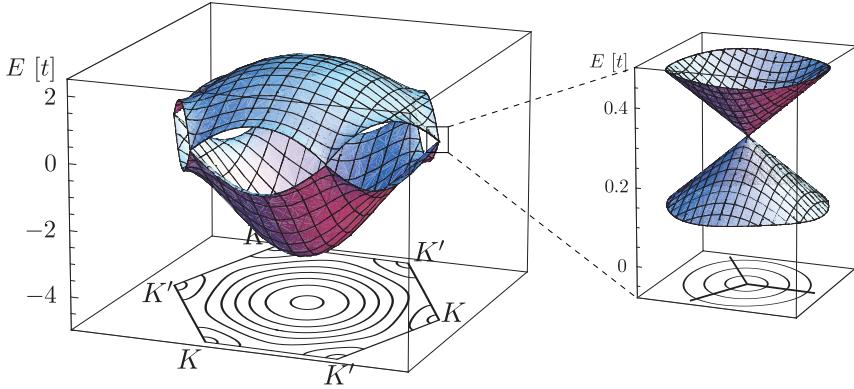


Figure G.3 – Left: Band structure of graphene in next-nearest neighbor approximation for $V_0 = M = 0$. Right: zoom-in of the band structure close to the K -point.

where $\sigma^{x,y,z}$ are the Pauli matrices. Here we can see that third-nearest neighbor hopping would only contribute terms of the form σ^x and σ^y , just like the nearest neighbor hopping. This is due to the fact that third-nearest neighbor hopping again connects atoms in different sublattices. In contrast, next-nearest neighbor hopping, connecting atoms of the same sublattice, introduces new terms in this Hamiltonian. In this sense, the inclusion of next-nearest neighbor hopping is more important than third-nearest neighbors.

This decomposition of the Hamiltonian in terms of Pauli matrices is just a formal paraphrase, as any 2×2 Hermitian matrix can be written as a sum of Pauli matrices and unity. Still, since this form is reminiscent of a Hamiltonian involving spin, the two-component character of the wave function $(\varphi_A, \varphi_B)^t$ due to the sublattice structure of graphene is commonly referred to as *pseudospin*. This analogy to spin will become even more apparent in the next section, when we consider a low-energy approximation to the Hamiltonian.

Having evaluated $H(\mathbf{k})$ we can now turn to solving the eigenproblem in Eq. (G.19). The eigenenergies are

$$E_{\pm}(\mathbf{k}) = V_0 - t' \left(2 \cos(k_x a) + 4 \cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{\sqrt{3}k_y a}{2}\right) \right) \pm \sqrt{M^2 + t'^2 \left(3 + 2 \cos(k_x a) + 4 \cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{\sqrt{3}k_y a}{2}\right) \right)}. \quad (\text{G.24})$$

We first consider the case where $V_0 = 0$ (a finite V_0 simply shifts the zero of energy) and $M = 0$. In Fig. G.3 we show $E_{\pm}(\mathbf{k})$ for \mathbf{k} in the first Brillouin zone of graphene.

The two bands $E_+(\mathbf{k})$ and $E_-(\mathbf{k})$ are always restricted to positive and negative energies, respectively, and are called the conduction and valence band. The conduction and the valence band touch at the six corners of the Brillouin zone for $E = 0$. Since the two atoms in the unit cell each contribute one electron, the valence band is completely filled. Thus, the Fermi energy is given as $E_F = 0$, and states at the Fermi energy have wave vector \mathbf{K} and \mathbf{K}' .

As seen from Fig. G.3, in our approximation the valence band has a larger band width than the conduction band. This is opposite to the usual band structure calculations for graphene, where the conduction band features a larger band width than the valence band [35, 190, 243]. This difference is due to the neglect of overlap from our side. However, for transport we are mainly concerned with states close to the Fermi energy, where overlap has only little effect³, and we will focus on this regime in the following.

Expanding the energy bands $E_\pm(\mathbf{k})$ to linear order around the two K -points, we obtain

$$E_\pm(\mathbf{K} + \mathbf{q}) = E_\pm(\mathbf{K}' + \mathbf{q}) \approx \hbar|\mathbf{q}|v_F, \quad (\text{G.25})$$

where $v_F = \sqrt{3}ta/2\hbar$ is the Fermi velocity. Thus, close to the K -points, the energy spectrum is linear, as shown in the zoom-in of Fig. G.3. In terms of band structure, the full energy cone is obtained by combining three 120° segments from the three equivalent corners of the Brillouin zone. Furthermore, t' does not enter in Eq. (G.25), as next-nearest neighbor hopping only introduces terms quadratic or higher in \mathbf{q} [35].

A linear energy spectrum is known from massless relativistic particles where $E = \pm\hbar kc$, and c is the speed of light. Close to the Fermi energy, the charge carriers in graphene thus have a “relativistic” energy dispersion, however with a much smaller “effective speed of light” $v_F \approx 10^6$ m/s. Because of this analogy, the cone-like shape of the energy bands around the K -points is commonly called *Dirac cone*. States within the Dirac cone at the K and K' -point are said to be in the K and K' -valley, respectively.

Finally, we consider the effect of a finite staggered potential. As seen from Eq. (G.24), the staggered potential opens a gap of size $2|M|$ at the corners of the Brillouin zone. (Fig. G.4). In this case, the spectrum close to the K -points can be approximated as

$$E_\pm(\mathbf{K} + \mathbf{q}) = E_\pm(\mathbf{K}' + \mathbf{q}) \approx \pm\sqrt{(M/v_F^2)^2 v_F^4 + (\hbar|\mathbf{q}|)^2 v_F^2}, \quad (\text{G.26})$$

which is analogous to the energy spectrum of a massive relativistic particle, $E = \pm\sqrt{m_0^2 c^4 + (\hbar k)^2 c^2}$, where m_0 is the rest mass of the particle. Hence, the staggered potential introduces a “mass” M/v_F^2 for the charge carriers in graphene.

Although the low-energy spectrum of graphene resembles relativistic energy dispersions, one should not erroneously conclude that this is due to relativistic effects in

³ When taking into account the overlap of neighboring orbitals, the Schrödinger equation in tight-binding approximation takes the form $H\varphi = ES\varphi$, where S is the overlap matrix. Roughly speaking, the smaller the energy E , the smaller the influence of the overlap, as E multiplies S . Thus, close to $E = 0$, the influence of overlap can be neglected. For example, this is reflected in Fig. 2 in Ref. [190], where the results of band structure calculations including and neglecting overlap are compared.

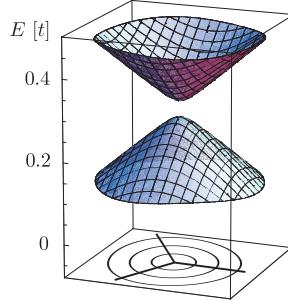


Figure G.4 – The staggered potential opens a gap in the band structure close to the K -points.

the motion of the *valence electrons* in the lattice of carbon atoms. Indeed, the starting point of the derivation was the non-relativistic Hamiltonian (G.5), and the linear low-energy spectrum is solely due to the lattice structure of the honeycomb network. Nevertheless, the *Bloch electrons* do behave in a way analogous to relativistic particles. In fact, this analogy goes even further than just a linear energy dispersion, as shown in the next section.

G.2.3 Effective Hamiltonian

The band structure of the previous section was obtained for a system with translational invariance. If this translational invariance is broken by, for example, a potential, it is in general not possible any more to solve the tight-binding problem analytically, and one has to resort to numerics or approximations. Here, we will derive such an approximation valid for slowly varying potentials and energies close to the Dirac point. This *effective Hamiltonian* for graphene was first derived by McClure including a magnetic field [244] and later by DiVincenzo and Mele including an external potential [245]. Both accounts followed the approach by Luttinger and Kohn for deriving an effective mass Hamiltonian [221], and so will we below.

In contrast to the lattice-periodic part of the Hamiltonian, a general potential may scatter Bloch electrons, connecting different Bloch vectors \mathbf{k} and \mathbf{k}' . Thus, we have to generalize the ansatz of Eq. (G.18) to include a range of wave vectors:

$$\varphi(\mathbf{x}) = \int d\mathbf{k} \varphi_A(\mathbf{k}) \Phi_A(\mathbf{k}, \mathbf{x}) + \varphi_B(\mathbf{k}) \Phi_B(\mathbf{k}, \mathbf{x}). \quad (\text{G.27})$$

Note that this expression is still exact, as the $\Phi_\alpha(\mathbf{k}, \mathbf{x})$ form a complete set within the tight-binding subspace. Assuming energies close to the Dirac point and a potential that varies slowly on the scale of the lattice, we can neglect scattering events between the K and K' point, as they are separated by a wave vector on the order of the reciprocal

lattice. Then, the ansatz (G.27) can be reduced to contain only wave vectors close to \mathbf{K} and \mathbf{K}' and we may write⁴

$$\begin{aligned}\varphi(\mathbf{x}) = & \int d\mathbf{q} \varphi_A(\mathbf{q}) \Phi_A(\mathbf{K} + \mathbf{q}, \mathbf{x}) + \varphi_B(\mathbf{q}) \Phi_B(\mathbf{K} + \mathbf{q}, \mathbf{x}) \\ & + \varphi'_A(\mathbf{q}) \Phi_A(\mathbf{K}' + \mathbf{q}, \mathbf{x}) + \varphi'_B(\mathbf{q}) \Phi_B(\mathbf{K}' + \mathbf{q}, \mathbf{x}).\end{aligned}\quad (\text{G.28})$$

When we use the ansatz (G.28) in Eq. (G.5), we can again make use of the fact that the tight-binding sums form an orthonormal set (see Eq. (G.17)). Thus, we need to calculate matrix elements of the form $\langle \Phi_\alpha(\mathbf{K}') | H | \Phi_\beta(\mathbf{K}') \rangle$. Since the lattice periodic part of the Hamiltonian does not mix different wave vectors, all corresponding matrix elements vanish for $\mathbf{q} \neq \mathbf{q}'$, and we obtain the result of the previous section, Eq. (G.23). However, for the external potential V_{ext} , we need to consider also matrix elements between different Bloch vectors. Making use of the smoothness of V_{ext} we can neglect matrix elements between different K -points, as was already done in the ansatz (G.28). Then, we find for the K -point

$$(E - H(\mathbf{K} + \mathbf{q})) \begin{pmatrix} \varphi_A(\mathbf{q}) \\ \varphi_B(\mathbf{q}) \end{pmatrix} + \int d\mathbf{q}' (V(\mathbf{q}, \mathbf{q}') + M(\mathbf{q}, \mathbf{q}') \sigma^z) \begin{pmatrix} \varphi_A(\mathbf{q}') \\ \varphi_B(\mathbf{q}') \end{pmatrix} = 0, \quad (\text{G.29})$$

where we again decomposed the external potential V_{ext} into a potential $V(\mathbf{x})$ that has the same value on sublattice A and B, and a staggered potential $M(\mathbf{x})$. $V(\mathbf{q}, \mathbf{q}')$ and $M(\mathbf{q}, \mathbf{q}')$ are defined as $\langle \Phi_A(\mathbf{K} + \mathbf{q}) | V | \Phi_A(\mathbf{K} + \mathbf{q}') \rangle$ and $\langle \Phi_A(\mathbf{K} + \mathbf{q}) | M | \Phi_A(\mathbf{K} + \mathbf{q}') \rangle$, respectively (These definitions only refer to sublattice A, but sublattice B is contained in the unit matrix and σ^z in Eq. (G.29)). The same equation holds for the K' -point, if we substitute \mathbf{K}' for \mathbf{K} .

These equations can be simplified further by means of two approximations: First, we restrict ourselves to the low-energy regime and expand $H(\mathbf{K} + \mathbf{q})$ and $H(\mathbf{K}' + \mathbf{q})$, keeping only terms up to linear order in \mathbf{q} . For this, we introduce the notation

$$\mathbf{K}_\tau = \begin{pmatrix} \tau \frac{4\pi}{3a} \\ 0 \end{pmatrix} \quad (\text{G.30})$$

such that we can treat both the K ($\tau = +1$) and the K' -point ($\tau = -1$) in a unified way. Then we find from Eq. (G.23)

$$H(\mathbf{K}_\tau + \mathbf{q}) \approx \hbar v_F \begin{pmatrix} 0 & \tau q_x - iq_y \\ \tau q_x + iq_y & 0 \end{pmatrix} = \hbar v_F (\tau q_x \sigma^x + q_y \sigma^y). \quad (\text{G.31})$$

Second, we again make use of the assumption that the potential $V(\mathbf{x})$ varies smoothly on the lattice scale, so that we can replace $V(\mathbf{q}, \mathbf{q}')$ by $V(\mathbf{q} - \mathbf{q}')$, where $V(\mathbf{q}) =$

⁴ Effective mass theory (also called $\mathbf{k} \cdot \mathbf{p}$ -theory) also uses often an ansatz different from Eq. (G.28): $\varphi(\mathbf{x}) \sim \int d\mathbf{q} \varphi(\mathbf{q}) e^{i\mathbf{q}\mathbf{x}} \Phi(\mathbf{K}, \mathbf{x})$. For example, this approach is used in the context of graphene in Refs. [244, 245]. As discussed in Ref. [221, Section II, III and Appendix A], these approaches are equivalent. However, the ansatz (G.28) seems more convenient for the present purposes.

$\frac{1}{(2\pi)^2} \int d\mathbf{x} e^{-i\mathbf{q}\mathbf{x}} V(\mathbf{x})$ is the Fourier transformation of $V(\mathbf{x})$, as discussed in Ref. [221].⁵ This also holds for the staggered potential $M(\mathbf{x})$.

With these approximations, Eq. (G.29) for the K -point reads

$$(E - \hbar v_F \mathbf{q} \cdot \boldsymbol{\sigma}) \begin{pmatrix} \varphi_A(\mathbf{q}) \\ \varphi_B(\mathbf{q}) \end{pmatrix} + \int d\mathbf{q}' (V(\mathbf{q} - \mathbf{q}') + M(\mathbf{q} - \mathbf{q}') \sigma^z) \begin{pmatrix} \varphi_A(\mathbf{q}') \\ \varphi_B(\mathbf{q}') \end{pmatrix} = 0. \quad (\text{G.32})$$

Introducing the *envelope wave functions* $\varphi_\alpha^{(l)}(\mathbf{x}) = \int d\mathbf{q} e^{i\mathbf{q}\mathbf{x}} \varphi_\alpha^{(l)}(\mathbf{q})$ and taking the Fourier transformation of Eq. (G.32), we finally arrive at

$$(E - H_K) \begin{pmatrix} \varphi_A(\mathbf{x}) \\ \varphi_B(\mathbf{x}) \end{pmatrix} = 0, \quad (\text{G.33})$$

where H_K is the effective Hamiltonian for graphene at the K -point,

$$H_K = v_F (-i\hbar \nabla) \cdot \boldsymbol{\sigma} + V(\mathbf{x}) + M(\mathbf{x}) \sigma^z. \quad (\text{G.34})$$

This effective Hamiltonian is identical to the Dirac Hamiltonian in two spatial dimensions with an effective speed of light v_F [246]. The potential $V(\mathbf{x})$ plays the role of a scalar potential, whereas the staggered potential $M(\mathbf{x})$ acts as a mass term.

In the previous section, we have pointed out that the low-energy dispersion of graphene resembles relativistic spectra. This relativistic dispersion is a consequence of the fact that the effective low-energy Hamiltonian of graphene is the Dirac Hamiltonian. However, the Dirac Hamiltonian leads to effects beyond just a relativistic spectrum: In particular, the two components of the wave function arising from the two sublattices A and B behave like a spinor under transformation. Hence, the notion of pseudospin for the sublattice degree of freedom is justified. Several of the unique properties of graphene are a consequence of pseudospin, such as the odd-integer quantum Hall effect [32] and Klein tunneling [34].

For the K' -point we obtain the effective Hamiltonian

$$H_{K'} = -v_F (-i\hbar \nabla) \cdot \boldsymbol{\sigma}^* + V(\mathbf{x}) + M(\mathbf{x}) \sigma^z. \quad (\text{G.35})$$

We can write the effective Hamiltonians for the K and K' -point in a unified way by defining

$$H_\tau = -i\hbar v_F (\tau \partial_x \sigma^x + \partial_y \sigma^y) + V(\mathbf{x}) + M(\mathbf{x}) \sigma^z, \quad (\text{G.36})$$

so that $H_K = H_+$ and $H_{K'} = H_-$. Note that the effective Hamiltonian does not contain any reference to the next-nearest neighbor hopping t' , as we only included terms linear in the momentum. Since t' only contributes starting from terms quadratic

⁵ There are some subtleties for the case of graphene: Graphene is a two-dimensional crystal, but the p_z -orbitals and the potentials also extend into the third dimension. A two-dimensional potential is then obtained by averaging the potential over the electron density in z -direction [245]. Here, we simply assume that such an averaging procedure has already been performed and assume $V(\mathbf{x})$ and $M(\mathbf{x})$ as two-dimensional potentials.

in \mathbf{q} , the influence of next-nearest neighbor hopping is expected to be negligible in the low-energy regime (with the notable exception of the graphene edge state, as discussed in Chapter 6).

Finally, we establish a connection between the envelope wave functions $\varphi_\alpha^{(t)}(\mathbf{x})$ and the tight-binding wave function $\varphi_\alpha(j)$. From Eq. (G.28) and the definition of the tight-binding sums (G.16) we find

$$\begin{aligned} \varphi(\mathbf{x}) &= \sum_{\alpha} \int d\mathbf{q} \sum_{\mathbf{R}_i} \left(e^{i(\mathbf{K}+\mathbf{q})(\mathbf{R}_i+\mathbf{d}_{\alpha})} \varphi_{\alpha}(\mathbf{q}) + e^{i(\mathbf{K}'+\mathbf{q})(\mathbf{R}_i+\mathbf{d}_{\alpha})} \varphi'_{\alpha}(\mathbf{q}) \right) \\ &\quad \times \chi_{2p_z}(\mathbf{x} - \mathbf{R}_i - \mathbf{d}_{\alpha}) \\ &= \sum_{\alpha} \sum_{\mathbf{R}_i} \left(e^{i\mathbf{K}(\mathbf{R}_i+\mathbf{d}_{\alpha})} \varphi_{\alpha}(\mathbf{R}_i + \mathbf{d}_{\alpha}) + e^{i\mathbf{K}'(\mathbf{R}_i+\mathbf{d}_{\alpha})} \varphi'_{\alpha}(\mathbf{R}_i + \mathbf{d}_{\alpha}) \right) \\ &\quad \times \chi_{2p_z}(\mathbf{x} - \mathbf{R}_i - \mathbf{d}_{\alpha}). \end{aligned} \quad (\text{G.37})$$

Comparing this expression to Eq. (G.10), we can identify

$$\varphi_{\alpha}(j) = e^{i\mathbf{K}(\mathbf{R}_j+\mathbf{d}_{\alpha})} \varphi_{\alpha}(\mathbf{R}_j + \mathbf{d}_{\alpha}) + e^{i\mathbf{K}'(\mathbf{R}_j+\mathbf{d}_{\alpha})} \varphi'_{\alpha}(\mathbf{R}_j + \mathbf{d}_{\alpha}). \quad (\text{G.38})$$

G.3 Graphene nanoribbons

G.3.1 Band structure

In this section we discuss how to understand the general features of the band structure of graphene nanoribbons from geometrical considerations and the band structure of an extended graphene sheet (for a similar discussion, see [169]).

A graphene nanoribbon (GNR) is a stripe cut out of the graphene honeycomb network. The boundary of the GNR depends on the direction of the cut. There are two basic types of GNRs: The GNR with zigzag edges (obtained by cutting e.g. in x -direction) and with armchair edges (obtained by cutting e.g. in y -direction) as shown in Figs. G.5(a) and (b), respectively. Since a GNR does not have the full translational symmetry of the graphene lattice, the electronic states of the GNR will in general be a linear combination of Bloch waves from the extended graphene sheet, with different wave vectors such that they satisfy the boundary condition of the nanoribbon.

We first consider the zigzag nanoribbon. The zigzag GNR has translational symmetry in x -direction, with a unit cell of length a , as indicated in Fig. G.5(a), and thus $k_x \in [-\pi/a, \pi/a]$ is a good quantum number. The (bulk) states of the zigzag GNR are then a linear combination of Bloch waves with a fixed value of k_x , as indicated in the sketch of the Brillouin zone in Fig. G.5(a). In particular, the different K -points project uniquely to the k_x -axis, with $k_x = -\frac{2\pi}{3a}$ for the K -point and $k_x = +\frac{2\pi}{3a}$ for the K' -point.⁶ Thus, the K and K' -points are decoupled in zigzag nanoribbons. In the

⁶ Note that this is not contradictory to our previous definition of $\mathbf{K} = (+\frac{4\pi}{3a}, 0)^t$ and $\mathbf{K}' = (-\frac{4\pi}{3a}, 0)^t$.

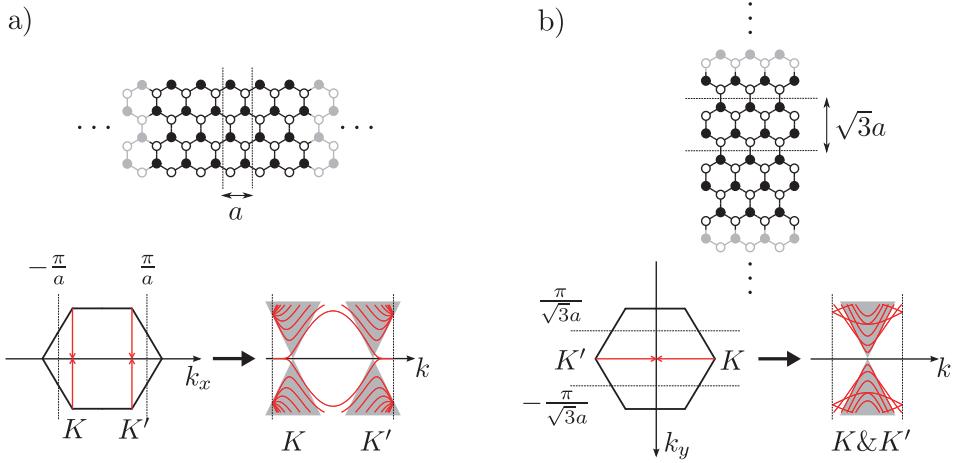


Figure G.5 – The general features of the band structure of (a) zigzag and (b) armchair nanoribbons can be deduced from geometrical considerations. The upper panels show the graphene nanoribbons and the corresponding unit cells. The lower panels show the orientation of the one-dimensional nanoribbon Brillouin zone within the two-dimensional graphene Brillouin zone and the nanoribbon band structure.

low energy regime of the zigzag GNR band structure, we observe the Dirac cones at these points (sketched in grey in the band structure of Fig. G.5(a)).

In addition to the states obtained as linear combinations of bulk Bloch waves, the zigzag GNR band structure also features states close to $E = 0$ for $k_x < -\frac{2\pi}{3a}$ and $k_x > \frac{2\pi}{3a}$. These are states located at the nanoribbon edge and cannot be expressed as a linear combination of bulk Bloch waves, and are discussed in detail in Chapter 6.

In contrast to a zigzag ribbon, the armchair ribbon has translational symmetry in y -direction, with a unit cell of length $\sqrt{3}a$. Then, $k_y \in [-\frac{\pi}{\sqrt{3}a}, \frac{\pi}{\sqrt{3}a}]$ is a good quantum number. The electronic states of the armchair nanoribbon are thus linear combinations of bulk Bloch waves with a fixed k_y . In particular, both the K and K' -point map to $k_y = 0$. Hence, the armchair edge mixes the K and K' -valley.

G.3.2 Boundary conditions

In order to handle finite-size graphene systems within the effective Dirac Hamiltonian, we need to derive appropriate boundary conditions. This will be done for the representative examples of zigzag and armchair nanoribbons.

The one-dimensional Brillouin zone of the nanoribbon has a smaller k_x -extent than the two-dimensional Brillouin zone of graphene, and $k_x = +\frac{4\pi}{3a}$ and $k_x = -\frac{2\pi}{3a}$ are connected by a one-dimensional reciprocal lattice vector of length $2\pi/a$.

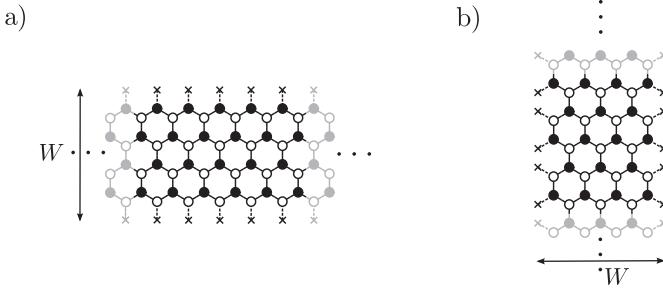


Figure G.6 – Boundary conditions for (a) zigzag and (b) armchair nanoribbons: The wave function must vanish on the first row of missing atoms (indicated by crosses).

Since the effective Hamiltonian does not carry any reference to the next-nearest neighbor hopping, it is enough to take into account nearest neighbors when deriving the boundary condition. At a graphene edge, we therefore demand that the wave function must vanish on the missing atoms that are adjacent to the edge.

In the case of the zigzag nanoribbon, Fig. G.6(a), the edge atoms are all from the *same* sublattice, and so is the first row of missing atoms. Thus, the boundary conditions are

$$e^{iK_x x} \varphi_A(x, 0) + e^{iK'_x x} \varphi'_A(x, 0) = 0, \quad (\text{G.39a})$$

$$e^{iK_x x} \varphi_B(x, W) + e^{iK'_x x} \varphi'_B(x, W) = 0, \quad (\text{G.39b})$$

where we used the connection between the envelope functions from the effective Hamiltonian and the tight-binding wave function (G.38). Here, we chose a coordinate system such that the first row of missing atoms is located at $y = 0$, and W is the distance to the first row of missing atoms on the other edge [53] (see Fig. G.6(a)). Since $K_x = -K'_x$ and Eqs. (G.39) must hold for any point x on the boundary, the envelope functions $\varphi_\alpha^{(r)}(\mathbf{x})$ must satisfy the boundary conditions separately for the zigzag ribbon:

$$\varphi_A(x, 0) = 0, \quad (\text{G.40a})$$

$$\varphi'_A(x, 0) = 0, \quad (\text{G.40b})$$

$$\varphi_B(x, W) = 0, \quad (\text{G.40c})$$

$$\varphi'_B(x, W) = 0. \quad (\text{G.40d})$$

In the case of the armchair ribbon, both types of sublattices are present at the

boundary, and hence the boundary condition reads

$$\varphi_A(x, 0) + \varphi'_A(x, 0) = 0, \quad (G.41a)$$

$$\varphi_B(x, 0) + \varphi'_B(x, 0) = 0, \quad (G.41b)$$

$$e^{iK_x W} \varphi_A(x, W) + e^{-iK_x W} \varphi'_A(x, W) = 0, \quad (G.41c)$$

$$e^{iK_x W} \varphi_B(x, W) + e^{-iK_x W} \varphi'_B(x, W) = 0. \quad (G.41d)$$

References

- [1] G. E. MOORE. *Cramming more components onto integrated circuits*. Electronics **38** (1965).
- [2] A. L. ROBINSON. *One billion transistors on a chip?*. Science **223**, 267–268 (1984).
- [3] S. THOMPSON, P. PACKAN, and M. BOHR. *MOS scaling: Transistor challenges for the 21st century*. Intel Tech. J. **Q3** (1998).
- [4] M. SCHULZ. *The end of the road for silicon?*. Nature **399**, 729–730 (1999).
- [5] P. A. PACKAN. *Pushing the limits*. Science **285**, 2079–2081 (1999).
- [6] P. S. PEERCY. *The drive to miniaturization*. Nature **406**, 1023–1026 (2000).
- [7] S. LLOYD. *Ultimate physical limits to computation*. Nature **406**, 1047–1054 (2000).
- [8] V. V. ZHIRNOV, R. K. CAVIN, and J. A. HUTCHBY. *Limits to binary logic switch scaling?A gedanken model*. Proc. IEEE **91**, 1934–1939 (2003).
- [9] J. D. MEINDL, Q. CHEN, and J. A. DAVIS. *Limits on silicon nanoelectronics for terascale integration*. Science **293**, 2044–2049 (2001).
- [10] M. LUNDSTROM. *Enhanced: Moore’s law forever?*. Science **299**, 210–211 (2003).
- [11] S. A. WOLF, D. D. AWSCHALOM, R. A. BUHRMAN, J. M. DAUGHTON, S. VON MOLNAR, M. L. ROUKES, A. Y. CHTCHELKANOVA, and D. M. TREGER. *Spintronics: A spin-based electronics vision for the future*. Science **294**, 1488–1495 (2001).
- [12] I. ŽUTIĆ, J. FABIAN, and S. DAS SARMA. *Spintronics: Fundamentals and applications*. Rev. Mod. Phys. **76**, 323–410 (2004).

- [13] J. FABIAN, A. MATOS-ABIAGUE, C. ERTLER, P. STANO, and I. ŽUTIĆ. *Semiconductor spintronics*. Acta Phys. Slovaca **4&5**, 565–907 (2007).
- [14] G. BINASCH, P. GRÜNBERG, F. SAURENBACH, and W. ZINN. *Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange*. Phys. Rev. B **39**, 4828–4830 (1989).
- [15] M. N. BAIBICH, J. M. BROTO, A. FERT, F. N. VAN DAU, F. PETROFF, P. EITENNE, G. CREUZET, A. FRIEDERICH, and J. CHAZELAS. *Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices*. Phys. Rev. Lett. **61**, 2472–2475 (1988).
- [16] M. JULLIÈRE. *Tunneling between ferromagnetic films*. Phys. Lett. A **54**, 225–226 (1975).
- [17] J. S. MOODERA, L. R. KINDER, T. M. WONG, and R. MESERVEY. *Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions*. Phys. Rev. Lett. **74**, 3273–3276 (1995).
- [18] T. MIYAZAKI and N. TEZUKA. *Giant magnetic tunneling effect in Fe/Al₂O₃/Fe junction*. J. Mag. Mag. Mater. **139**, L231–L234 (1995).
- [19] C. CHAPPERT, A. FERT, and F. N. VAN DAU. *The emergence of spin electronics in data storage*. Nat. Mater. **6**, 813–823 (2007).
- [20] S. DATTA and B. DAS. *Electronic analog of the electro-optic modulator*. Appl. Phys. Lett. **56**, 665–667 (1990).
- [21] Y. A. BYCHKOV and E. I. RASHBA. *Oscillatory effects and the magnetic susceptibility of carriers in inversion layers*. J. Phys. C: Solid State Phys. **17**, 6039–6045 (1984).
- [22] Y. A. BYCHKOV and E. I. RASHBA. *Properties of a 2d electron gas with lifted spectral degeneracy*. JETP Lett. **39**, 78–81 (1984).
- [23] J. SCHLIEMANN, J. C. EGUES, and D. LOSS. *Nonballistic spin-field-effect transistor*. Phys. Rev. Lett. **90**, 146801 (2003).
- [24] S. BANDYOPADHYAY and M. CAHAY. *Alternate spintronic analog of the electro-optic modulator*. Appl. Phys. Lett. **85**, 1814–1816 (2004).
- [25] M. I. D'YAKONOV and V. I. PEREL'. *Possibility of orienting electron spins*. JETP Lett. **13**, 467–469 (1971).
- [26] J. E. HIRSCH. *Spin Hall effect*. Phys. Rev. Lett. **83**, 1834–1837 (1999).
- [27] S. MURAKAMI, N. NAGAOSA, and S.-C. ZHANG. *Dissipationless quantum spin current at room temperature*. Science **301**, 1348–1351 (2003).

- [28] J. SINOVA, D. CULCER, Q. NIU, N. A. SINITSYN, T. JUNGWIRTH, and A. H. MACDONALD. *Universal intrinsic spin Hall effect*. Phys. Rev. Lett. **92**, 126603 (2004).
- [29] K. S. NOVOSELOV, A. K. GEIM, S. V. MOROZOV, D. JIANG, Y. ZHANG, S. V. DUBONOS, I. V. GRIGORIEVA, and A. A. FIRSOV. *Electric field effect in atomically thin carbon films*. Science **306**, 666–669 (2004).
- [30] N. D. MERMIN. *Crystalline order in two dimensions*. Phys. Rev. **176**, 250–254 (1968).
- [31] A. K. GEIM and K. S. NOVOSELOV. *The rise of graphene*. Nat. Mater. **6**, 183–191 (2007).
- [32] K. S. NOVOSELOV, A. K. GEIM, S. V. MOROZOV, D. JIANG, M. I. KATSNELSON, I. V. GRIGORIEVA, S. V. DUBONOS, and A. A. FIRSOV. *Two-dimensional gas of massless dirac fermions in graphene*. Nature **438**, 197–200 (2005).
- [33] K. S. NOVOSELOV, Z. JIANG, Y. ZHANG, S. V. MOROZOV, H. L. STORMER, U. ZEITLER, J. C. MAAN, G. S. BOEBINGER, P. KIM, and A. K. GEIM. *Room-temperature quantum Hall effect in graphene*. Science **315**, 1379 (2007).
- [34] M. I. KATSNELSON, K. S. NOVOSELOV, and A. K. GEIM. *Chiral tunnelling and the Klein paradox in graphene*. Nat. Phys. **2**, 620–625 (2006).
- [35] A. H. CASTRO NETO, F. GUINEA, N. M. R. PERES, K. S. NOVOSELOV, and A. K. GEIM. *The electronic properties of graphene*. To be published in Rev. Mod. Phys., arXiv:0709.1163v2 (2008).
- [36] K. BOLOTIN, K. SIKES, Z. JIANG, M. KLIMA, G. FUDENBERG, J. HONE, P. KIM, and H. STORMER. *Ultrahigh electron mobility in suspended graphene*. Solid State Commun. **146**, 351–355 (2008).
- [37] K. I. BOLOTIN, K. J. SIKES, J. HONE, H. L. STORMER, and P. KIM. *Temperature dependent transport in suspended graphene*. arXiv:0805.1830v1 (2008).
- [38] B. TRAUZETTEL, D. V. BULAEV, D. LOSS, and G. BURKARD. *Spin qubits in graphene quantum dots*. Nature Phys. **3**, 192–196 (2007).
- [39] D. HUERTAS-HERNANDO, F. GUINEA, and A. BRATAAS. *Spin-orbit coupling in curved graphene, fullerenes, nanotubes, and nanotube caps*. Phys. Rev. B **74**, 155426 (2006).
- [40] H. MIN, J. E. HILL, N. A. SINITSYN, B. R. SAHU, L. KLEINMAN, and A. H. MACDONALD. *Intrinsic and Rashba spin-orbit interactions in graphene sheets*. Phys. Rev. B **74**, 165310 (2006).

- [41] D. HUERTAS-HERNANDO, F. GUINEA, and A. BRATAAS. *Spin relaxation times in disordered graphene*. Eur. Phys. J. Spec. Top. **148**, 177–181 (2007).
- [42] E. W. HILL, A. K. GEIM, K. S. NOVOSELOV, F. SCHEDIN, and P. BLAKE. *Graphene spin valve devices*. IEEE Trans. Magn. **42**, 2694–2696 (2006).
- [43] N. TOMBROS, C. JOZSA, M. POPINCIUC, H. T. JONKMAN, and B. J. VAN WEES. *Electronic spin transport and spin precession in single graphene layers at room temperature*. Nature **448**, 571–574 (2007).
- [44] S. CHO, Y.-F. CHEN, and M. S. FUHRER. *Gate-tunable graphene spin valve*. Appl. Phys. Lett. **91**, 123105 (2007).
- [45] N. TOMBROS, S. TANABE, A. VELIGURA, C. JOZSA, M. POPINCIUC, H. T. JONKMAN, and B. J. VAN WEES. *Anisotropic spin relaxation in graphene*. Phys. Rev. Lett. **101**, 046601 (2008).
- [46] C. JÓZSA, M. POPINCIUC, N. TOMBROS, H. T. JONKMAN, and B. J. VAN WEES. *Electronic spin drift in graphene field-effect transistors*. Phys. Rev. Lett. **100**, 236603 (2008).
- [47] J. MOSER, A. MATOS-ABIAGUE, D. SCHUH, W. WEGSCHEIDER, J. FABIAN, and D. WEISS. *Tunneling anisotropic magnetoresistance and spin-orbit coupling in Fe/GaAs/Au tunnel junctions*. Phys. Rev. Lett. **99**, 056601 (2007).
- [48] M. LOBENHOFER. Unpublished.
- [49] D. A. GRIER. *Human computers: the first pioneers of the information age*. Endvr. **25**, 28–32 (2001).
- [50] J. WURM, A. RYCERZ, I. ADAGIDELI, M. WIMMER, K. RICHTER, and H. U. BARANGER. *Symmetry classes in graphene quantum dots: Universal spectral statistics, weak localization, and conductance fluctuations*. arXiv:0808.1008v1 (2008).
- [51] M. SCHEID, M. WIMMER, D. BERCIOUX, and K. RICHTER. *Zeeman ratchets for ballistic spin currents*. Phys. Status Solidi (c) **3**, 4235–4238 (2006).
- [52] I. ADAGIDELI, M. SCHEID, M. WIMMER, G. E. W. BAUER, and K. RICHTER. *Extracting current-induced spins: spin boundary conditions at narrow Hall contacts*. New J. Phys. **9**, 382 (2007).
- [53] J. WURM. *Quantum Transport in Graphene-Based Nanosystems*. Diplomarbeit, Universität Regensburg (2008).
- [54] M. SCHEID, A. PFUND, D. BERCIOUX, and K. RICHTER. *Coherent spin ratchets: A spin-orbit based quantum ratchet mechanism for spin-polarized currents in ballistic conductors*. Phys. Rev. B **76**, 195303 (2007).

- [55] M. SCHEID, D. BERCIOUX, and K. RICHTER. *Zeeman ratchets: pure spin current generation in mesoscopic conductors with non-uniform magnetic fields.* New J. Phys. **9**, 401 (2007).
- [56] V. KRÜCKL. *Mesoskopischer Transport in zweidimensionalen Lochgasen.* Diplomarbeit, Universität Regensburg (2008).
- [57] M. WIMMER. Unpublished.
- [58] A. L. FETTER and J. WALECKA. *Quantum Theory of Many-Particle Systems* (Dover, Mineola, 2003).
- [59] S. DATTA. *Electronic Transport in Mesoscopic Transport* (Cambridge University Press, Cambridge, 2002).
- [60] D. K. FERRY and S. M. GOODNICK. *Transport in Nanostructures* (Cambridge University Press, Cambridge, 2001).
- [61] H. HAUG and A.-P. JAUHO. *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer, Berlin, Heidelberg, 1998).
- [62] H. BRUUS and K. FLENSBERG. *Many-body Quantum Theory in Condensed Matter Physics: An Introduction* (Oxford University Press, Oxford, 2004).
- [63] J. RAMMER. *Quantum Field Theory of Non-equilibrium States* (Cambridge University Press, Cambridge, 2007).
- [64] L. P. KADANOFF and G. BAYM. *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
- [65] L. V. KELDYSH. *Diagram technique for nonequilibrium processes.* Sov. Phys. JETP **20**, 1018–1026 (1965).
- [66] J. RAMMER and H. SMITH. *Quantum field-theoretical methods in transport theory of metals.* Rev. Mod. Phys. **58**, 323–359 (1986).
- [67] P. DANIELEWICZ. *Quantum theory of nonequilibrium processes, I.* Ann. Phys. (NY) **152**, 239–304 (1984).
- [68] M. WAGNER. *Expansions of nonequilibrium Green's functions.* Phys. Rev. B **44**, 6104–6117 (1991).
- [69] S. HERSHFIELD. *Reformulation of steady state nonequilibrium quantum statistical mechanics.* Phys. Rev. Lett. **70**, 2134–2137 (1993).
- [70] F. SCHWABL. *Advanced Quantum Mechanics* (Springer, Berlin, 1999).

- [71] R. A. CRAIG. *Perturbation expansion for real-time Green's functions.* J. Math. Phys. **9**, 605–611 (1968).
- [72] D. C. LANGRETH. *Linear and nonlinear response theory with applications.* In: *Linear and Nonlinear Electron Transport*, edited by J. T. PEVREESE and E. VAN DOREN, pp. 3–32 (Plenum, New York, 1976).
- [73] S. DATTA. *Electrical resistance: an atomistic view.* Nanotech. **15**, S433–S451 (2004).
- [74] G. D. MAHAN. *Many-Particle Physics* (Kluwer, New York, NY, 2005).
- [75] D. C. LANGRETH and J. W. WILKINS. *Theory of spin resonance in dilute magnetic alloys.* Phys. Rev. B **6**, 3189–3227 (1972).
- [76] J. H. DAVIES, S. HERSHFIELD, P. HYLDGAARD, and J. W. WILKINS. *Current and rate equation for resonant tunneling.* Phys. Rev. B **47**, 4603–4618 (1993).
- [77] G. E. KIMBALL and G. H. SHORTLEY. *The numerical solution of Schrödinger's equation.* Phys. Rev. **45**, 815–820 (1934).
- [78] L. PAULING and E. B. WILSON. *Introduction to Quantum Mechanics* (Dover, New York, 1935).
- [79] D. FRUSTAGLIA, M. HENTSCHEL, and K. RICHTER. *Aharonov-Bohm physics with spin. II. Spin-flip effects in two-dimensional ballistic systems.* Phys. Rev. B **69**, 155327 (2004).
- [80] P. HAVU, V. HAVU, M. J. PUSKA, and R. M. NIEMINEN. *Nonequilibrium electron transport in two-dimensional nanostructures modeled using Green's functions and the finite-element method.* Phys. Rev. B **69**, 115325 (2004).
- [81] R. C. BOWEN, G. KLIMECK, R. K. LAKE, W. R. FRENSLEY, and T. MOISE. *Quantitative simulation of a resonant tunneling diode.* J. Appl. Phys. **81**, 3207–3213 (1997).
- [82] 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.* Phys. Rev. B **59**, 11936–11948 (1999).
- [83] M. LUISIER, A. SCHENK, W. FICHTNER, and G. KLIMECK. *Atomistic simulation of nanowires in the $sp^3d^5s^*$ tight-binding formalism: From boundary conditions to strain calculations.* Phys. Rev. B **74**, 205323 (2006).
- [84] M. BRANDBYGE, J.-L. MOZOS, P. ORDEJÓN, J. TAYLOR, and K. STOKBRO. *Density-functional method for nonequilibrium electron transport.* Phys. Rev. B **65**, 165401 (2002).

- [85] A. DI CARLO, A. PECCIA, L. LATESSA, T. FRAUENHEIM, and G. SEIFERT. *Tight-binding DFT for molecular electronics (gDFTB)*. In: *Introducing Molecular Electronics*, edited by G. CUNIBERTI, G. FAGAS, and K. RICHTER, pp. 153–184 (Springer, Berlin, Heidelberg, 2006).
- [86] A. R. ROCHA, V. M. GARCÍA-SUÁREZ, S. BAILEY, C. LAMBERT, J. FERRER, and S. SANVITO. *Spin and molecular electronics in atomically generated orbital landscapes*. Phys. Rev. B **73**, 085414 (2006).
- [87] Y. MEIR and N. S. WINGREEN. *Landauer formula for the current through an interacting electron region*. Phys. Rev. Lett. **68**, 2512–2515 (1992).
- [88] R. LANDAUER. *Spatial variation of currents and fields due to localized scatterers in metallic conduction*. IBM J. Res. Dev. **1**, 223–231 (1957).
- [89] M. BÜTTIKER, Y. IMRY, R. LANDAUER, and S. PINHAS. *Generalized many-channel conductance formula with application to small rings*. Phys. Rev. B **31**, 6207–6215 (1985).
- [90] D. S. FISHER and P. A. LEE. *Relation between conductivity and transmission matrix*. Phys. Rev. B **23**, 6851–6854 (1981).
- [91] A. LASSL. *Spin-dependent Transport of Interacting Electrons in Mesoscopic Systems*. Dissertation, Universität Regensburg (2007).
- [92] T. USUKI, M. TAKATSU, R. A. KIEHL, and N. YOKOYAMA. *Numerical analysis of electron-wave detection by a wedge-shaped point contact*. Phys. Rev. B **50**, 7615–7625 (1994).
- [93] N. W. ASHCROFT and N. D. MERMIN. *Solid State Physics* (Harcourt College Publishers, Orlando, FL, 1976).
- [94] R. E. ALLEN. *Green's function and generalized phase shift for surface and interface problems*. Phys. Rev. B **19**, 917–924 (1979).
- [95] R. E. ALLEN. *Green's functions for surface physics*. Phys. Rev. B **20**, 1454–1472 (1979).
- [96] D. H. LEE and J. D. JOANNOPOULOS. *Simple scheme for surface-band calculations. I*. Phys. Rev. B **23**, 4988–4996 (1981).
- [97] D. H. LEE and J. D. JOANNOPOULOS. *Simple scheme for surface-band calculations. II. The Green's function*. Phys. Rev. B **23**, 4997–5004 (1981).
- [98] M. P. LOPEZ SANCHO, J. M. LOPEZ SANCHO, and J. RUBIO. *Quick iterative scheme for the calculation of transfer matrices: application to Mo (100)*. J. Phys. F: Met. Phys. **14**, 1205–1215 (1984).

- [99] M. P. LOPEZ SANCHO, J. M. LOPEZ SANCHO, and J. RUBIO. *Highly convergent schemes for the calculation of bulk and surface Green functions.* J. Phys. F: Met. Phys. **15**, 851–858 (1985).
- [100] J. VELEV and W. BUTLER. *On the equivalence of different techniques for evaluating the Green function for a semi-infinite system using a localized basis.* J. Phys.: Condens. Matter **16**, R637–R657 (2004).
- [101] A. UMERSKI. *Closed-form solutions to surface Green's functions.* Phys. Rev. B **55**, 5266–5275 (1997).
- [102] P. S. KRSTIĆ, X.-G. ZHANG, and W. H. BUTLER. *Generalized conductance formula for the multiband tight-binding model.* Phys. Rev. B **66**, 205319 (2002).
- [103] A. MACKINNON. *The calculation of transport properties and density of states of disordered solids.* Z. Phys. B **59**, 385–390 (1985).
- [104] T. ANDO. *Quantum point contacts in magnetic fields.* Phys. Rev. B **44**, 8017–8027 (1991).
- [105] P. A. KHOMYAKOV, G. BROCKS, V. KARPAN, M. ZWIERZYCKI, and P. J. KELLY. *Conductance calculations for quantum wires and interfaces: Mode matching and Green's functions.* Phys. Rev. B **72**, 035450 (2005).
- [106] F. TISSEUR and K. MEERBERGEN. *The quadratic eigenvalue problem.* SIAM Rev. **43**, 235–286 (2001).
- [107] E. ANDERSON, Z. BAI, C. BISCHOF, S. BLACKFORD, J. DEMMEL, J. DONAGARRA, J. DU CROZ, A. GREENBAUM, S. HAMMARLING, A. MCKENNEY, and D. SORENSEN. *LAPACK Users' Guide*, 3rd edition (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).
- [108] T. B. BOYKIN. *Generalized eigenproblem method for surface and interface states: The complex bands of GaAs and AlAs.* Phys. Rev. B **54**, 8107–8115 (1996).
- [109] R. P. FEYNMAN. *Forces in molecules.* Phys. Rev. **56**, 340–343 (1939).
- [110] H. HELLMANN. *Einführung in die Quantenchemie* (Deuticke, Leipzig, 1937).
- [111] C. W. J. BEENAKKER and H. VAN HOUTEN. *Quantum transport in semiconductor nanostructures.* Solid State Phys. **44**, 1–228 (1991).
- [112] G. H. GOLUB and C. F. VAN LOAN. *Matrix Computations*, 3rd edition (The John Hopkins University Press, 1996).
- [113] A. R. WILLIAMS, P. J. FEIBELMAN, and N. D. LANG. *Green's-function methods for electronic-structure calculations.* Phys. Rev. B **26**, 5433–5444 (1982).

- [114] T. USUKI, M. SAITO, M. TAKATSU, R. A. KIEHL, and N. YOKOYAMA. *Numerical analysis of ballistic-electron transport in magnetic fields by using a quantum point contact and a quantum wire.* Phys. Rev. B **52**, 8244–8255 (1995).
- [115] C. J. LAMBERT and D. WEAIRE. *Decimation and Anderson localization.* Phys. Status Solidi (b) **101**, 591–595 (1980).
- [116] M. LEADBEATER and C. J. LAMBERT. *A decimation method for studying transport properties of disordered systems.* Ann. Phys. **7**, 498–502 (1998).
- [117] D. MAMALUY, D. VASILESKA, M. SABATHIL, T. ZIBOLD, and P. VOGL. *Contact block reduction method for ballistic transport and carrier densities of open nanostructures.* Phys. Rev. B **71**, 245321 (2005).
- [118] D. J. THOULESS and S. KIRKPATRICK. *Conductivity of the disordered linear chain.* J. Phys. C: Solid State Phys. **14**, 235–245 (1981).
- [119] P. A. LEE and D. S. FISHER. *Anderson localization in two dimensions.* Phys. Rev. Lett. **47**, 882–885 (1981).
- [120] H. U. BARANGER, D. P. DiVINCENZO, R. A. JALABERT, and A. D. STONE. *Classical and quantum ballistic-transport anomalies in microjunctions.* Phys. Rev. B **44**, 10637–10675 (1991).
- [121] R. LAKE, G. KLIMECK, R. C. BOWEN, and D. JOVANOVIC. *Single and multi-band modeling of quantum electron transport through layered semiconductor devices.* J. Appl. Phys. **81**, 7845–7869 (1997).
- [122] A. LASSL, P. SCHLAGHECK, and K. RICHTER. *Effects of short-range interactions on transport through quantum point contacts: A numerical approach.* Phys. Rev. B **75**, 045346 (2007).
- [123] P. DROUVELIS, P. SCHMELCHER, and P. BASTIAN. *Parallel implementation of the recursive Green's function method.* J. Comp. Phys. **215**, 741–756 (2006).
- [124] T. KRAMER, E. J. HELLER, and R. E. PARROTT. *An efficient and accurate method to obtain the energy-dependent Green function for general potentials.* J. Phys.: Conf. Ser. **99**, 012010 (2008).
- [125] S. ROTTER, J.-Z. TANG, L. WIRTZ, J. TROST, and J. BURGDÖRFER. *Modular recursive Green's function method for ballistic quantum transport.* Phys. Rev. B **62**, 1950–1960 (2000).
- [126] S. ROTTER, B. WEINGARTNER, N. ROHRINGER, and J. BURGDÖRFER. *Ballistic quantum transport at high energies and high magnetic fields.* Phys. Rev. B **68**, 165302 (2003).

- [127] N. E. GIBBS, J. WILLIAM G. POOLE, and P. K. STOCKMEYER. *An algorithm for reducing the bandwidth and profile of a sparse matrix*. SIAM J. Num. Anal. **13**, 236–250 (1976).
- [128] E. CUTHILL and J. MCKEE. *Reducing the bandwidth of sparse symmetric matrices*. In: *Proceedings of the 1969 24th national conference*, pp. 157–172 (ACM, New York, 1969).
- [129] A. GEORGE. *Computer implementation of the finite element method*. Technical Report STAN-CS-71-208, Computer Sci. Dept., Stanford Univ., Stanford, CA (1971).
- [130] W.-H. LIU and A. H. SHERMAN. *Comparative analysis of the Cuthill–McKee and the reverse Cuthill–McKee ordering algorithms for sparse matrices*. SIAM J. Num. Anal. **13**, 198–213 (1976).
- [131] B. KERNIGHAN and S. LIN. *An efficient heuristic procedure for partitioning graphs*. Bell Syst. Tech. J. **49**, 291–308 (1970).
- [132] C. M. FIDUCCIA and R. M. MATTHEYES. *A linear-time heuristic for improving network partitions*. In: *DAC '82: Proceedings of the 19th conference on Design automation*, pp. 175–181 (IEEE Press, Piscataway, NJ, USA, 1982).
- [133] G. KARYPIS and V. KUMAR. *Multilevel k-way hypergraph partitioning*. VLSI Design, **11**, 285–300 (2000).
- [134] B. HENDRICKSON and E. ROTHBERG. *Improving the run time and quality of nested dissection ordering*. SIAM J. Sci. Comput. **20**, 468–489 (1998).
- [135] J. O'NEIL and D. B. SZYLD. *A block ordering method for sparse matrices*. SIAM J. Sci. Stat. Comput. **11**, 811–823 (1990).
- [136] A. COON and M. STADTHERR. *Generalized block-tridiagonal matrix orderings for parallel computation in process flowsheeting*. Comput. Chem. Eng. **19**, 787–805 (1995).
- [137] K. V. CAMARDA and M. A. STADTHERR. *Matrix ordering strategies for process engineering: graph partitioning algorithms for parallel computation*. Comput. Chem. Eng. **23**, 1063–1073 (1999).
- [138] B. HENDRICKSON. *Graph partitioning and parallel solvers: Has the emperor no clothes?*. In: *Workshop on Parallel Algorithms for Irregularly Structured Problems*, pp. 218–225 (1998).
- [139] B. HENDRICKSON and T. G. KOLDA. *Partitioning rectangular and structurally unsymmetric sparse matrices for parallel processing*. SIAM J. Sci. Comput. **21**, 2048–2072 (2000).

- [140] C. AYKANAT, A. PINAR, and ÜMIT V. ÇATALYÜREK. *Permuting sparse rectangular matrices into block-diagonal form*. SIAM J. Sci. Comput. **25**, 1860–1879 (2004).
- [141] R. SEDGEWICK. *Algorithms in C++*, 2nd edition (Addison Wesley, 1992).
- [142] A. GUPTA. *Fast and effective algorithms for graph partitioning and sparse-matrix ordering*. IBM J. Res. Dev. **41**, 171–184 (1997).
- [143] H. D. SIMON and S.-H. TENG. *How good is recursive bisection?*. SIAM J. Sci. Comput. **18**, 1436–1445 (1997).
- [144] M. R. GAREY and D. S. JOHNSON. *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W. H. Freeman & Co., New York, NY, USA, 1990).
- [145] R. C. WHALEY, A. PETITET, and J. J. DONGARRA. *Automated empirical optimization of software and the ATLAS project*. Parallel Comput. **27**, 3–35 (2001).
- [146] K. KAZYMYRENKO and X. WAINTEL. *Knitting algorithm for calculating Green functions in quantum systems*. Phys. Rev. B **77**, 115119 (2008).
- [147] Z. H. QIAO and J. WANG. *A variant transfer matrix method suitable for transport through multi-probe systems*. Nanotech. **18**, 435402 (2007).
- [148] K. v. KLITZING, G. DORDA, and M. PEPPER. *New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance*. Phys. Rev. Lett. **45**, 494–497 (1980).
- [149] C. GOULD, C. RÜSTER, T. JUNGWIRTH, E. GIRGIS, G. M. SCHOTT, R. GI-RAUD, K. BRUNNER, G. SCHMIDT, and L. W. MOLENKAMP. *Tunneling anisotropic magnetoresistance: A spin-valve-like tunnel magnetoresistance using a single magnetic layer*. Phys. Rev. Lett. **93**, 117203 (2004).
- [150] M. LOBENHOFER. *TAMR-Effekt in epitaktischen Fe/GaAs-Systemen*. Diplomarbeit, Universität Regensburg (2007).
- [151] A. MATOS-ABIAGUE and J. FABIAN. *Spin-orbit induced anisotropy in the tunneling magnetoresistance of magnetic tunnel junctions*. arXiv:cond-mat/0702387v2 (2007).
- [152] A. MATOS-ABIAGUE and J. FABIAN. *Anisotropic tunneling magnetoresistance and tunneling anisotropic magnetoresistance: spin-orbit coupling in magnetic tunnel junctions*. arXiv:0807.3564v1 (2008).

- [153] C. COHEN-TANNOUDJI, B. DIU, and F. LALOË. *Qunatum Mechanics* (John Wiley & Sons, New York, NY, 1977).
- [154] H. U. BARANGER and A. D. STONE. *Electrical linear-response theory in an arbitrary magnetic field: A new Fermi-surface formation.* Phys. Rev. B **40**, 8169–8193 (1989).
- [155] M. WIMMER. *Spinpolarisierter Transport in Ferromagnet/Halbleiter-Tunnelsystemen.* Diplomarbeit, Universität Regensburg (2004).
- [156] L. EAVES, K. W. H. STEVENS, and F. W. SHEARD. *Tunnel currents and electron tunnelling times in semiconductor heterostructure barriers in the presence of an applied magnetic field.* In: *The physics and fabrication of microstructures and microdevices*, edited by M. KELLY and C. WEISBUCH, volume 13 of *Springer Proceedings in Physics* (Springer, Berlin, 1986).
- [157] M. I. KATSNELSON. *Graphene: carbon in two dimensions.* Mater. Today **10**, 20–27 (2007).
- [158] M. Y. HAN, B. ÖZYILMAZ, Y. ZHANG, and P. KIM. *Energy band-gap engineering of graphene nanoribbons.* Phys. Rev. Lett. **98**, 206805 (2007).
- [159] Z. CHEN, Y.-M. LIN, M. J. ROOKS, and P. AVOURIS. *Graphene nano-ribbon electronics.* Physica E **40**, 228–232 (2007).
- [160] X. LI, X. WANG, L. ZHANG, S. LEE, and H. DAI. *Chemically derived, ultra-smooth graphene nanoribbon semiconductors.* Science **319**, 1229–1232 (2008).
- [161] X. WANG, Y. OUYANG, X. LI, H. WANG, J. GUO, and H. DAI. *Room temperature all semiconducting sub-10nm graphene nanoribbon field-effect transistors.* arXiv:0803.3464v1, to appear in Phys. Rev. Lett. (2008).
- [162] I. MARTIN and Y. M. BLANTER. *Transport in disordered graphene nanoribbons.* arXiv:0705.0532v2 (2007).
- [163] F. SOLS, F. GUINEA, and A. H. C. NETO. *Coulomb blockade in graphene nanoribbons.* Phys. Rev. Lett. **99**, 166803 (2007).
- [164] Y.-W. SON, M. L. COHEN, and S. G. LOUIE. *Energy gaps in graphene nanoribbons.* Phys. Rev. Lett. **97**, 216803 (2006).
- [165] M. EVALDSSON, I. V. ZOZOULENKO, H. XU, and T. HEINZEL. *Edge disorder induced Anderson localization and conduction gap in graphene nanoribbons.* arXiv:0805.4326v1 (2008).
- [166] E. R. MUCCIOLO, A. H. C. NETO, and C. H. LEWENKOPF. *Conductance quantization and transport gap in disordered graphene nanoribbons.* arXiv:0806.3777v1 (2008).

- [167] L. TAPASZTO, G. DOBRIK, P. LAMBIN, and L. P. BIRO. *Tailoring the atomic structure of graphene nanoribbons by scanning tunnelling microscope lithography.* Nature Nanotech. **3**, 397–401 (2008).
- [168] S. S. DATTA, D. R. STRACHAN, S. M. KHAMIS, and A. T. C. JOHNSON. *Crystallographic etching of few-layer graphene.* Nano Lett. **8**, 1912–1915 (2008).
- [169] K. NAKADA, M. FUJITA, G. DRESSELHAUS, and M. S. DRESSELHAUS. *Edge state in graphene ribbons: Nanometer size effect and edge shape dependence.* Phys. Rev. B **54**, 17954–17961 (1996).
- [170] M. FUJITA, K. WAKABAYASHI, K. NAKADA, and K. KUSAKABE. *Peculiar localized state at zigzag graphite edge.* J. Phys. Soc. Jpn. **65**, 1920–1923 (1996).
- [171] A. R. AKHMEROV and C. W. J. BEENAKKER. *Boundary conditions for Dirac fermions on a terminated honeycomb lattice.* Phys. Rev. B **77**, 085423 (2008).
- [172] Y. NIIMI, T. MATSUI, H. KAMBARA, K. TAGAMI, M. TSUKADA, and H. FUKUYAMA. *Scanning tunneling microscopy and spectroscopy studies of graphite edges.* App. Surf. Sci. **241**, 43–48 (2005).
- [173] S. BANERJEE, M. SARDAR, N. GAYATHRI, A. K. TYAGI, and B. RAJ. *Conductivity landscape of highly oriented pyrolytic graphite surfaces containing ribbons and edges.* Phys. Rev. B **72**, 075418 (2005).
- [174] Y. NIIMI, T. MATSUI, H. KAMBARA, K. TAGAMI, M. TSUKADA, and H. FUKUYAMA. *Scanning tunneling microscopy and spectroscopy of the electronic local density of states of graphite surfaces near monoatomic step edges.* Phys. Rev. B **73**, 085421 (2006).
- [175] S. OKADA and A. OSHIYAMA. *Magnetic ordering in hexagonally bonded sheets with first-row elements.* Phys. Rev. Lett. **87**, 146803 (2001).
- [176] Y.-W. SON, M. L. COHEN, and S. G. LOUIE. *Half-metallic graphene nanoribbons.* Nature **444**, 347–349 (2006).
- [177] F. MUÑOZ ROJAS, D. JACOB, J. FERNÁNDEZ-ROSSIER, and J. J. PALACIOS. *Coherent transport in graphene nanoconstrictions.* Phys. Rev. B **74**, 195417 (2006).
- [178] T. C. LI and S.-P. LU. *Quantum conductance of graphene nanoribbons with edge defects.* Phys. Rev. B **77**, 085408 (2008).
- [179] L. P. ZÂRBO and B. K. NIKOLIC. *Spatial distribution of local currents of massless Dirac fermions in quantum transport through graphene nanoribbons.* Europhys. Lett. **80**, 47001 (2007).

- [180] D. ARESHKIN, D. GUNLYCKE, and C. WHITE. *Ballistic transport in graphene nanostrips in the presence of disorder: Importance of edge effects.* Nano Lett. **7**, 204–210 (2007).
- [181] K. WAKABAYASHI, Y. TAKANE, and M. SIGRIST. *Perfectly conducting channel and universality crossover in disordered graphene nanoribbons.* Phys. Rev. Lett. **99**, 036601 (2007).
- [182] N. M. R. PERES, F. GUINEA, and A. H. C. NETO. *Electronic properties of disordered two-dimensional carbon.* Phys. Rev. B **73**, 125411 (2006).
- [183] K. SASAKI, S. MURAKAMI, and R. SAITO. *Stabilization mechanism of edge states in graphene.* Appl. Phys. Lett. **88**, 113110 (2006).
- [184] D. POREZAG, T. FRAUENHEIM, T. KÖHLER, G. SEIFERT, and R. KASCHNER. *Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon.* Phys. Rev. B **51**, 12947–12957 (1995).
- [185] L. BREY and H. A. FERTIG. *Electronic states of graphene nanoribbons studied with the Dirac equation.* Phys. Rev. B **73**, 235411 (2006).
- [186] M. V. BERRY and R. J. MONDRAGON. *Neutrino billiards: time-reversal symmetry-breaking without magnetic fields.* Proc. R. Soc. Lond. A **412**, 53–74 (1987).
- [187] E.-N. FOO, M. F. THORPE, and D. WEAIRE. *Effective surface potential method for calculating surface states.* Surf. Sci. **57**, 323–347 (1976).
- [188] A. R. AKHMEROV, J. H. BARDARSON, A. RYCERZ, and C. W. J. BEENAKKER. *Theory of the valley-valve effect in graphene nanoribbons.* Phys. Rev. B **77**, 205416 (2008).
- [189] A. RYCERZ, J. TWORZYDŁO, and C. W. J. BEENAKKER. *Valley filter and valley valve in graphene.* Nat. Phys. **3**, 172–175 (2007).
- [190] S. REICH, J. MAULTZSCH, C. THOMSEN, and P. ORDEJÓN. *Tight-binding description of graphene.* Phys. Rev. B **66**, 035412 (2002).
- [191] Y. YAO, F. YE, X.-L. QI, S.-C. ZHANG, and Z. FANG. *Spin-orbit gap of graphene: First-principles calculations.* Phys. Rev. B **75**, 041401 (2007).
- [192] 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.* Phys. Rev. B **62**, R4790–R4793 (2000).
- [193] E. I. RASHBA. *Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem.* Phys. Rev. B **62**, R16267–R16270 (2000).

- [194] A. FERT and H. JAFFRÈS. *Conditions for efficient spin injection from a ferromagnetic metal into a semiconductor.* Phys. Rev. B **64**, 184420 (2001).
- [195] M. WIMMER, I. ADAGIDELI, S. BERBER, D. TOMÁNEK, and K. RICHTER. *Spin currents in rough graphene nanoribbons: Universal fluctuations and spin injection.* Phys. Rev. Lett. **100**, 177207 (2008).
- [196] S. BERBER. Unpublished.
- [197] L. PISANI, J. A. CHAN, B. MONTANARI, and N. M. HARRISON. *Electronic structure and magnetic properties of graphitic ribbons.* Phys. Rev. B **75**, 064418 (2007).
- [198] E. H. LIEB. *Two theorems on the hubbard model.* Phys. Rev. Lett. **62**, 1201–1204 (1989).
- [199] E. C. STONER. *Collective electron ferromagnetism. II. Energy and specific heat.* Proc. R. Soc. London. Ser. A **169**, 339–371 (1939).
- [200] N. D. MERMIN and H. WAGNER. *Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models.* Phys. Rev. Lett. **17**, 1133–1136 (1966).
- [201] O. V. YAZYEV and M. I. KATSNELSON. *Magnetic correlations at graphene edges: Basis for novel spintronics devices.* Phys. Rev. Lett. **100**, 047209 (2008).
- [202] D.-E. JIANG, B. G. SUMPTER, and S. DAI. *First principles study of magnetism in nanographenes.* J. Chem. Phys. **127**, 124703 (2007).
- [203] B. HUANG, F. LIU, J. WU, B.-L. GU, and W. DUAN. *Suppression of spin polarization in graphene nanoribbons by edge defects and impurities.* Phys. Rev. B **77**, 153411 (2008).
- [204] S. BHOWMICK and V. B. SHENOY. *Edge state magnetism of single layer graphene nanostructures.* J. Chem. Phys. **128**, 244717 (2008).
- [205] O. HOD, V. BARONE, and G. E. SCUSERIA. *Half-metallic graphene nanodots: A comprehensive first-principles theoretical study.* Phys. Rev. B **77**, 035411 (2008).
- [206] P. W. ANDERSON, D. J. THOUESS, E. ABRAHAMS, and D. S. FISHER. *New method for a scaling theory of localization.* Phys. Rev. B **22**, 3519–3526 (1980).
- [207] C. W. J. BEENAKKER. *Random-matrix theory of quantum transport.* Rev. Mod. Phys. **69**, 731–808 (1997).
- [208] O. N. DOROKHOV. *Transmission coefficient and the localization length of an electron in N bound disordered chains.* JETP Lett. **36**, 318–321 (1982).

- [209] P. A. MELLO, P. PEREYRA, and N. KUMAR. *Macroscopic approach to multi-channel disordered conductors*. Ann. Phys. (N.Y.) **181**, 290–317 (1988).
- [210] M. E. GERTSENSHTEIN and V. B. VASIL’EV. *Waveguides with random inhomogeneities and brownian motion in the lobachevsky plane*. Theory Probab. Appl. **4**, 391–398 (1959).
- [211] C. W. J. BEENAKKER and J. A. MELSEN. *Conductance fluctuations, weak localization, and shot noise for a ballistic constriction in a disordered wire*. Phys. Rev. B **50**, 2450–2457 (1994).
- [212] J. REPP. Private communication.
- [213] I. ADAGIDELI, J. H. BARDARSON, and P. JACQUOD. *Electrical probing of the spin conductance of mesoscopic cavities*. arXiv:0808.2756v1 (2008).
- [214] I. ADAGIDELI, G. E. W. BAUER, and B. I. HALPERIN. *Detection of current-induced spins by ferromagnetic contacts*. Phys. Rev. Lett. **97**, 256601 (2006).
- [215] M. JOHNSON and R. H. SILSBEY. *Coupling of electronic charge and spin at a ferromagnetic-paramagnetic metal interface*. Phys. Rev. B **37**, 5312–5325 (1988).
- [216] J. TWORZYDŁO, B. TRAUZETTEL, M. TITOV, A. RYCERZ, and C. W. J. BEENAKKER. *Sub-poissonian shot noise in graphene*. Phys. Rev. Lett. **96**, 246802 (2006).
- [217] D. XIAO, W. YAO, and Q. NIU. *Valley-contrasting physics in graphene: Magnetic moment and topological transport*. Phys. Rev. Lett. **99**, 236809 (2007).
- [218] J. SCHLIEMANN. *Spin Hall effect*. Int. J. Mod. Phys. B **20**, 1015–1036 (2006).
- [219] B. K. NIKOLIĆ, L. P. ZÂRBO, and S. SOUMA. *Mesoscopic spin Hall effect in multiprobe ballistic spin-orbit-coupled semiconductor bridges*. Phys. Rev. B **72**, 075361 (2005).
- [220] J. H. BARDARSON, I. ADAGIDELI, and P. JACQUOD. *Mesoscopic spin Hall effect*. Phys. Rev. Lett. **98**, 196601 (2007).
- [221] J. M. LUTTINGER and W. KOHN. *Motion of electrons and holes in perturbed periodic fields*. Phys. Rev. **97**, 869–883 (1955).
- [222] M. HEAD-GORDON, P. E. MASLEN, and C. A. WHITE. *A tensor formulation of many-electron theory in a nonorthogonal single-particle basis*. J. Chem. Phys. **108**, 616–625 (1998).
- [223] L. A. PONOMARENKO, F. SCHEDIN, M. I. KATSNELSON, R. YANG, E. W. HILL, K. S. NOVOSELOV, and A. K. GEIM. *Chaotic Dirac billiard in graphene quantum dots*. Science **320**, 356–358 (2008).

- [224] E. MCCANN, K. KECHEDZHI, V. I. FAL'KO, H. SUZURA, T. ANDO, and B. L. ALTSCHULER. *Weak-localization magnetoresistance and valley symmetry in graphene*. Physical Review Letters **97**, 146805 (2006).
- [225] S. V. MOROZOV, K. S. NOVOSELOV, M. I. KATSNELSON, F. SCHEDIN, L. A. PONOMARENKO, D. JIANG, and A. K. GEIM. *Strong suppression of weak localization in graphene*. Phys. Rev. Lett. **97**, 016801 (2006).
- [226] X. WU, X. LI, Z. SONG, C. BERGER, and W. A. DE HEER. *Weak antilocalization in epitaxial graphene: Evidence for chiral electrons*. Phys. Rev. Lett. **98**, 136801 (2007).
- [227] F. V. TIKHONENKO, D. W. HORSELL, R. V. GORBACHEV, and A. K. SAVCHENKO. *Weak localization in graphene flakes*. Phys. Rev. Lett. **100**, 056802 (2008).
- [228] I. ADAGIDELI. Private communication.
- [229] F. SCHWABL. *Quantum Mechanics* (Springer, Berlin, 2003).
- [230] A. CRESTI, R. FARCHIONI, G. GROSSO, and G. P. PARRAVICINI. *Keldysh-Green function formalism for current profiles in mesoscopic systems*. Phys. Rev. B **68**, 075306 (2003).
- [231] A. D. STONE and A. SZAFER. *What is measured when you measure a resistance? - the Landauer formula revisited*. IBM J. Res. Dev. **32**, 384–413 (1988).
- [232] A. GARCIA. *Numerical Methods for Physics* (Prentice Hall, Upper Saddle River, NJ, 1999).
- [233] R. PEIERLS. *Zur Theorie des Diamagnetismus von Leitungselektronen*. Z. f. Phys. A **80**, 763–791 (1933).
- [234] T. L. LI and K. J. KUHN. *Effects of mass discontinuity on the numerical solutions to quantum wells using the effective mass equation*. J. Comp. Phys. **110**, 292–300 (1994).
- [235] D. J. BEN-DANIEL and C. B. DUKE. *Space-charge effects on electron tunneling*. Phys. Rev. **152**, 683–692 (1966).
- [236] G. DRESSELHAUS. *Spin-orbit coupling effects in zinc blende structures*. Phys. Rev. **100**, 580–586 (1955).
- [237] U. RÖSSLER. *Solid State Theory* (Springer, Berlin, Heidelberg, New York, 2004).
- [238] P. R. WALLACE. *The band theory of graphite*. Phys. Rev. **71**, 622–634 (1947).

- [239] J. C. SLATER and G. F. KOSTER. *Simplified LCAO method for the periodic potential problem*. Phys. Rev. **94**, 1498–1524 (1954).
- [240] D. FINKENSTADT, G. PENNINGTON, and M. J. MEHL. *From graphene to graphite: A general tight-binding approach for nanoribbon carrier transport*. Phys. Rev. B **76**, 121405 (2007).
- [241] H. RAZA and E. KAN. *An extended Hückel theory based atomistic model for graphene nanoelectronics*. To appear in J. Comput. Electron. (2008).
- [242] F. D. M. HALDANE. *Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the "parity anomaly"*. Phys. Rev. Lett. **61**, 2015–2018 (1988).
- [243] R. SAITO, G. DRESSELHAUS, and M. S. DRESSELHAUS. *Physical Properties of Carbon Nanotubes* (Imperial College Press, London, 1998).
- [244] J. W. MCCLURE. *Diamagnetism of graphite*. Phys. Rev. **104**, 666–671 (1956).
- [245] D. P. DiVINCENZO and E. J. MELE. *Self-consistent effective-mass theory for intralayer screening in graphite intercalation compounds*. Phys. Rev. B **29**, 1685–1694 (1984).
- [246] M. E. PESKIN and D. V. SCHROEDER. *An Introduction to Quantum Field Theory* (Perseus Books, Cambridge, MA, 1995).

Acknowledgements

All's well that ends well

William Shakespeare (1564–1616)

A dissertation is the result of many intellectual struggles over several years. I was fortunate to be able to work in a group with an extremely enjoyable and stimulating working atmosphere. Now, I would like to use the opportunity to thank all the people that I benefitted from in one way or the other.

Above all, I would like to express my sincere gratitude to Prof. Dr. Klaus Richter for giving me the opportunity to work in his group. Over the course of this thesis, he has always been willing to listen to any problem and to provide helpful suggestions and comments. Without his guidance and patience, this thesis would not exist in this form.

In addition, I am also greatly indebted to Dr. İnanç Adagideli who introduced me to the fascinating physics of graphene. Discussions with him were often demanding, but always fruitful, and his ideas were a constant source of inspiration. Thank you, İnanç!

I would also like to thank the rest of the “graphene gang”, Jürgen Wurm and Dr. Adam Rycerz, for many stimulating discussions. The cooperation in this group has certainly been the most productive I have seen so far.

Furthermore, I would like to express my gratitude to Matthias Scheid who was always willing to share his spintronics knowledge with me.

The work on the TAMR effect was inspired through a cooperation of experimental and theoretical groups in Regensburg. In particular, I would like to thank Michael Lobenhofer for sharing his experimental data before publication, as well as Dr. Alex Matos-Abiague and Prof. Dr. Fabian for discussions about their TAMR-related work.

Many thanks also to Prof. Dr. Peter Spellucci for discussing numerical aspects of linear algebra with me on `sci.math.num-analysis`.

I am very grateful to all the people who were willing to proofread drafts in various stages of this thesis: Jürgen Wurm, Michael Hartung, Dr. İnanç Adagideli, Dr. Adam Rycerz, Matthias Scheid, Viktor Krückl and Jan Bundesmann. Any remaining mistakes are entirely within my own responsibility.

Over the past years, I was sharing a room with other colleagues. I was fortunate to meet officemates that always created a pleasant working atmosphere. My deepest thanks go to Dr. Tobias Paul, Dr. Andreas Laßl, Dr. Bernhard Rieder, Dr. Josef Kainz, Dr. Christian Lechner, Michael Hartung, Matthias Scheid, Dr. Arseni Goussev, and Timo Hartmann.

During my task as a system administrator for the chair I have worked with many other colleagues that always have been motivated and very helpful. Special thanks therefore to Dr. Christian Lechner, Dr. Dieter Hierl, Michael Hartung, and Dr. Christian Hagen. In this regard I would also like to especially mention Dr. Fritz Wünsch who always supported us in every possible way, and from whom we obtained a good insight into the work flow of university matters.

With respect to the whole complex systems group, I would like to thank every member who attended and kept the “Kaffee- und Teerunde” alive.

Last, but not least in the list of people I got to know in the university, I am very grateful to Angie Reisser, who was always helpful and encouraging—in virtually everything. The pleasant atmosphere in the group is to a good part her merit.

I acknowledge computing time at the Leibniz-Rechenzentrum München and financial support from the DFG within GRK 638 and SFB 689, and financial travel support from Verein der Freunde der Universität Regensburg.

Am Ende dieser Arbeit möchte ich noch ganz besonders meinen Eltern Christina und Herbert als auch meinem Bruder Christian für all die Unterstützung danken, die ich von ihnen erhalten habe. Ganz besonders gilt mein Dank auch meiner Frau Marlies: Danke, dass Du da warst—und sein wirst.