On the numerical computation of electron transport through a topological crystalline insulator

Bart de Leeuw

February 11, 2015

Masterscriptie Mathematical Sciences & Theoretical Physics

Supervised by Jason Frank and Lars Fritz

Universiteit Utrecht

Abstract

In this thesis we aim to compute the conductance of a disordered non-standard topological insulator numerically. We test several algorithms for this purpose and argue that the recursive Green's function algorithm is convenient for our purposes. We test the stability of the conductance under disorder of a specific topological crystalline insulator and find that in this case crystalline symmetries do not suffice to make the boundary modes robust.

Contents

Chapter 1

Introduction

In the last years topological insulators have become an important research topic in theoretical condensed matter physics [1, 2, 3]. These systems have the interesting properties that, if they have boundaries, electrons can propagate along the boundaries, while the bulk of the system is insulating. Another remarkable property is that the boundary conductance is not hindered by disorder in the system. This can be explained as electrons moving along a "one-way road", such that they can not backscatter when they stumble upon disorder. These topological phases arise due to symmetries of the system; in [1, 2, 3] topological phases of systems with only anti-unitary symmetries have been completely classified.

These anti-unitary symmetries do not tell the entire story however. Because topological insulators are usually realised in crystalline systems, we can also analyse the effect of crystalline symmetries on the existence of boundary modes. It turns out boundary modes can indeed exist [4, 5, 6, 7, 8], but a complete theory for these topological crystalline insulators where also the stability with respect to disorder is addressed does not yet exist.

The aim of this thesis is to study the stability under disorder of one such topological crystalline insulator. The model for this insulator is taken from [5] and the results have also been published separately in [9]. To study this system we combine numerical simulations with semi-analytical tools.

The numerical simulations are carried out by perturbing the system with disorder and then using the non-equilibrium Green's functions formalism [10, 11], which is used to express the conductance in terms of the Green's function of an operator depending on the system Hamiltonian and the way electrons are entered into and exit from the system.[12, 13] The simulations are then performed by computing conductances for various disorder realizations and parameters and comparing them.

In order to apply the non-equilibrium formalism it is important to specify the way electrons enter and exit the system (through *leads*, which model long wires) and to reduce the size of the operator to be inverted to the size of the matrix which specifies the system Hamiltonian. This can be done using the concept of the lead self-energy [13], which can be computed numerically [14].

When the inverse Green's function has been defined, the conductance can be obtained from a block of the Green's function. This means we should compute a part of the inverse of a matrix. This matrix to be partially inverted is mostly sparse, complex and non-hermitian and usually large. Another aim of this thesis is to describe and test algorithms to obtain the relevant part of the Green's function.

This thesis is structured as follows: in chapter 2 we give a short review of important concept for the transport of electrons through crystals, the effects disorder usually has and how the computation of the conductance can be interpreted as a scattering problem. Appendix A gives some more background information about the quantum mechanics used in this chapter.

In chapter 3 we describe the non-equilibrium Green's function formalism, introducing the lead self-energy and expressing the conductance in terms of Green's functions. This chapter also contains some simple examples of conductance computations and a general algorithm for the computation of the lead self-energy. Some more examples of conductance computations for simple disordered systems are given in appendix B, which also serves as a further illustration of concepts discussed in chapter 2.

In chapter 4 we compare various ways in which we could obtain the conductance numerically. This chapter is a combination of literature reviews and practical tests for algorithms that could potentially be used to compute conductances efficiently. Some background information on the numerical linear algebra used in this chapter can be found in appendix C.

Topological insulators are the topic of chapter 5. We first give a short introduction on topological insulators using an example. Afterwards we state our results on topological crystalline insulators: this part of the thesis is a slightly extended and reformulated version of [9]. We end this thesis by drawing some conclusions in chapter 6.

Chapter 2

Introduction to electron transport theory

In this chapter we introduce concepts from electron transport theory. We argue that electron transport is most easily described in crystals. We then introduce some models for electron transport in crystals and make a connection between those different models. Also discussed is the effect of impurities in the crystal on electron transport, since real-world crystals always contain some impurities.

Having introduced the crystal background on which the electrons move, we introduce the Landauer-Büttiker formalism to connect the movement of electrons through a system consisting of *leads* (wires that guide the electrons to the insulator we want to study) and an insulator of interest to physical properties of this insulator, such as conductance. It will turn out to be useful to describe this as a scattering problem, in which electrons are scattered from one lead into another by scattering through the insulator.

2.1 Electrons in crystals

We would like to describe the behaviour of electrons in solids. Since solids consist of many individual atoms or molecules that are relatively close together, each of those atoms or molecules contains many electrons and everything is described by the laws of quantum mechanics, this is in general not possible. If we however make some assumptions on the structure of the solid and the electrons therein, the problem can become a lot more tractable.

We assume that our atoms or molecules together with most of their electrons form a regular structure, a crystal, that influences how the rest of the electrons behave. We will also assume that any interactions between the non-bound electrons can be neglected due to screening by the atoms and

molecules with their bound electrons.

Depending on how we describe the interactions between bound matter and unbound electrons, we can use different models for electron behaviour in crystals. Here we will discuss two types of such models, the nearly-free electron model and tight-binding models. First however, we make more precise what we mean with a crystal and introduce some concepts from crystallography. A much more detailed introduction on crystals can be found in for example [15].

2.1.1 What is a crystal?

A crystal is a solid that is built up from large numbers of identical atoms or molecules, such that the positions of these building blocks are regular. With regular we mean that for an infinite extension of the crystal there exist discrete symmetries. For ordinary crystals, translations by vectors known as the *lattice* vectors are among those discrete symmetries. The points at which atoms or molecules are located are known as *lattice points*.

A common and convenient way to describe a crystal is by specifying a *basis cell* and the lattice vectors. A basis cell is defined as a part of the crystal such that the entire crystal can be constructed by taking the basis cell together with copies of the basis cell to which a shift by a lattice vector is applied any number of times. The smallest basis cell (which is in general not unique) is known as a *primitive* cell. A special kind of primitive cell is called *Wigner-Seitz* cell. This is a different name for the Voronoi cell [16] of any lattice point.

Since (perfect) crystals are periodic, we can apply a spatial Fourier transformation to them. This will again yield a lattice, known as the reciprocal lattice. The Voronoi cell around the origin of the reciprocal lattice is known as the *first Brillouin zone*.

2.1.2 Bloch's theorem

We will now consider a model for electron transport known as the nearly-free electron model. In this model, the electrons are described with quantum mechanical wave functions $\Psi_n(\mathbf{x})$ satisfying the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m_e}\Delta\Psi_n(\mathbf{x}) + V(\mathbf{x})\Psi_n(\mathbf{x}) = E_n\Psi_n(\mathbf{x}), \tag{2.1.2.1}$$

where $n \in \mathbb{N}$ is an index[1], $\mathbf{x}$ is a position vector, m_e the electron mass, Δ the Laplacian, $V(\mathbf{x})$ a periodic potential energy function with the same pe-

[1]Throughout this thesis, we use the convention that $0 \in \mathbb{N}$. We denote $\mathbb{N}\backslash\{0\}$ by $\mathbb{N}_+$.

riodicity as the crystal, E_n the energy corresponding to the n-th eigenstate and $\hbar := \frac{h}{2\pi}$ and h Planck's constant.

Bloch's theorem [17] now states that the solutions to (2.1.2.1) can be written as the product of a plane wave and a periodic function.

Theorem 1 (Bloch's theorem)**.** *A basis of solutions to (2.1.2.1) is given by* $\Psi_n(\mathbf{x}) = e^{i\mathbf{kx}}u_{n,\mathbf{k}}(\mathbf{x})$*, where* $\mathbf{k}$ *is a vector known as the crystal wave vector or Bloch momentum (taking values in the first Brillouin zone),* $u_{n,\mathbf{k}}(\mathbf{x})$ *are functions which have periodicities equal to the crystal lattice vectors and the index* n *is referred to as the band index.*

Proof. This proof is taken from [18]. We first show that any state that is a simultaneous eigenstate of the operators that translate the system along a lattice vector is a Bloch wave.

Suppose we have such a state $\Psi_n(\mathbf{x})$ for some n. Then $\Psi_n(\mathbf{x}+\mathbf{a_i}) = e^{2\pi i\theta_i}\Psi_n(\mathbf{x})$, where $\mathbf{a_i}$ is a lattice translation vector and $\theta_i \in \mathbb{R}$ (because we are on a crystal, the translation can not change the normalisation of Ψ but only a phase). Define $\mathbf{k} := \sum_{i=1}^{3}\theta_i\mathbf{b_i}$ and $u_{n,\mathbf{k}}(\mathbf{x}) := e^{-i\mathbf{kx}}\Psi_n(\mathbf{x})$, where $\mathbf{b_i}$ is the reciprocal lattice vector corresponding to $\mathbf{a_i}$. Then $u_{n,\mathbf{k}}(\mathbf{x}+\mathbf{a_i}) = e^{-i\mathbf{k}(\mathbf{x}+\mathbf{a_i})}\Psi(n,\mathbf{x}+\mathbf{a_i}) = \left(e^{-i\mathbf{kx}}e^{-2\pi i\theta_i}\right)\left(e^{2\pi i\theta_i}\Psi(n,\mathbf{x})\right) = u_{n,\mathbf{k}}(\mathbf{x})$. Hence $u_{n,\mathbf{k}}(\mathbf{x})$ has the same periodicity as the lattice and $\Psi(n,\mathbf{x}) = e^{i\mathbf{kx}}u_{n,\mathbf{k}}(\mathbf{x})$ is a Bloch wave.

We now note that any lattice translation vector commutes with the crystal Hamiltonian and any other lattice translation vector. This means (cf. for example appendix A or [19]) that there exists a simultaneous eigenbasis of the Hamiltonian and all translation vectors. By the arguments above, the elements of this eigenbasis are Bloch waves and since they are eigenfunctions of the Hamiltonian they are energy eigenstates solving (2.1.2.1). □

2.1.2.1 Band structure

Following [19], we now argue that the energy E as a function of the crystal wave vector $\mathbf{k}$ is continuous and periodic over the Brillouin zone. We do this by considering a one dimensional example, for which we will find one *energy band*. Afterwards we will generalize this to a system with multiple bands.

We first consider a one-dimensional system with some periodic potential and send the barrier height of this potential to ∞ over intervals of non-zero length. An infinite set of degenerate ground states of this system is given by states that are localised around one specific minimum of the potential (between two neighbouring infinite barriers). We can index these states by $|l\rangle$, $l \in \mathbb{Z}$.

The energy eigenstates $|l\rangle$ are not eigenstates of the translation operator, but we can find a family of eigenstates $|\theta\rangle := \sum_{l=-\infty}^{\infty} e^{il\theta} |l\rangle$, which are eigenstates of both the Hamiltonian and the translation operator for any $\theta \in [-\pi, \pi]$ [19].

If we now consider the situation where the barriers between potential minima are high but finite, we can define $\Delta_m := \langle l+m| \hat{H} |l\rangle$, where $\hat{H}$ is the system Hamiltonian and $m \in \mathbb{Z}$. Since the barriers are high, we assume Δ_m to go to zero rapidly as $m \to \infty$. We act with this $\hat{H}$ on $|\theta\rangle$

$$\begin{aligned}
\hat{H} \sum_{l=-\infty}^{\infty} e^{il\theta} |l\rangle &= \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} e^{il\theta} \Delta_m |l+m\rangle \\
&= \sum_{m=-\infty}^{\infty} \Delta_m \sum_{l=-\infty}^{\infty} e^{i(l-m)\theta} |l\rangle \\
&= \left(\sum_{m=-\infty}^{\infty} \Delta_m e^{-im\theta} \right) \sum_{l=-\infty}^{\infty} e^{il\theta} |l\rangle .
\end{aligned}$$

We recognize that the eigenvalue is a Fourier series and identify $\theta = ka$, where a is the lattice spacing and k the wave vector. The energy as a function of the Bloch momentum is now smooth and periodic over the Brillouin zone.

In this example, the entire spectrum is indexed only by one variable $k \in [-\pi/a, \pi/a]$ and we can define a smooth periodic map form $[-\pi/a, \pi/a]$ to $\mathbb{R}$ by $k \to \sum_{m=-\infty}^{\infty} \Delta_m e^{-im\theta}$. This map is known as (energy) band [20].

We have however seen in Bloch's theorem above that in general the Bloch eigenfunctions $u_{n,\mathbf{k}}$ are characterised not only by the Bloch momentum $\mathbf{k}$, but also by an index n.[2] We say such a system has n bands, since for each value of the index n we can define a smooth periodic map from the Brillouin zone to the energy eigenvalues corresponding to the Bloch eigenfunctions [20]. All bands together are known as the band structure of the system.

Band structure can also be used to explain the existence of conductors, semi-conductors and insulators. Suppose we have a system with multiple bands that are separated by (one or more) gaps.[3] There will be electrons filling all states available to them, starting from the one with the lowest

[2] In our simple 1D example above $\mathbf{k}$ becomes a scalar and n can take only 1 value.

[3] A gap can be defined as an interval of energies such that there exist lower and higher energies that are in the spectrum of the Hamiltonian $\hat{H}$, while this interval does not contain any eigenvalues of the $\hat{H}$, so these energies are not in any band, while above and below it there are bands. If no gaps exist, the system will always be a metal.

energy.[4]

At zero temperature, all states with energies below the chemical potential (see section 2.1.3) will be filled and all states with energies above it are empty. If the chemical potential is within some band, we can easily increase or decrease the number of electrons by varying the potential a little bit and (if there is no disorder, see section 2.1.4) electrons can move rather freely through the system, which is then known as *conductor*. If the chemical potential is in one of the gaps[5], the above process does not work and we are dealing with an *insulator*.[6]. The highest filled band of an insulator is called *valence band*, while the lowest empty band is the *conduction band*.

At non-zero temperature the electrons are distributed according to a probability distribution known as Fermi-Dirac distribution instead, but qualitatively the distinction between insulators and conductors remains the same.

2.1.2.2 Effective mass

The group velocity of a wave packet in band n is given by $\mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}} E(\mathbf{k})$. If we define the crystal momentum $\mathbf{p} = \hbar\mathbf{k}$, we can use Newton's second law $\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F} = m\mathbf{a}$, where t denotes time, $\mathbf{F}$ is a force, $\mathbf{a}$ an acceleration and m an effective mass (to be computed). Combining these formulas we get

$$\frac{1}{m_n} = \mathbf{a_n}\frac{1}{\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{p}}\mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar^2}\Delta_{\mathbf{k}} E(\mathbf{k}). \tag{2.1.2.2}$$

We remark that the effective mass can be either positive or negative, as opposed to "real" mass, which is always positive! Near the band edge (which is the region where we are most interested in) we can use a quadratic approximation of the band dispersion relation

$$E(\mathbf{k}) \approx E_n + \frac{\hbar^2\mathbf{k}^2}{2m_n}.$$

We can then smooth out the wave function and eliminate the periodicity from the problem by introducing the effective mass equation [13]

$$\left(E_n + \frac{(i\hbar\nabla + e\mathbf{A})^2}{2m_n} + U(\mathbf{x})\right)\Psi(\mathbf{x}) = E\Psi(\mathbf{x}), \tag{2.1.2.3}$$

where e is the charge of an electron, $\mathbf{A}$ is the vector potential (to couple the electron to an external electromagnetic field, see for example [19] or

[4] At zero temperature they will fill precisely those states with lowest energy, while there will be a different distribution at finite temperature, but low energy-states will always be preferred over high energy-states.

[5] or precisely at the maximum of a band

[6] Unless the gap is narrow, in which case we have a semi-conductor

[13]), $U(\mathbf{x})$ some potential energy (in general not periodic, since the effective mass takes care of the periodic lattice potential) and, with a slight abuse of notation, $\Psi(\mathbf{x})$ the smoothend-out wave function. We can see that (2.1.2.3) indeed induces the quadratic dispersion relation $E(\mathbf{k}) = E_n + \frac{\hbar^2 \mathbf{k}^2}{2m_n}$.

2.1.3 Tight-binding models

In addition to the nearly-free electron model, there are also other models for electron transport through crystals. One group of models is known as the *tight-binding models*. In these models the moving electrons are assumed to be localised at a certain site and they are only influenced by sites that are sufficiently close to them.

We will again be interested in finding eigenstates of the Hamiltonian operator $\mathcal{H}$, but in this case the Hamiltonian will not be a differential operator any more. We will obtain an expression for $\mathcal{H}$ using the formalism of second quantisation. If we assume the number of states M to be finite, we can write $\mathcal{H}$ as the product of a row vector $\mathbf{c}^\dagger$, a Hamiltonian matrix $\hat{H}$ and a column vector $\mathbf{c}$

$$\mathcal{H} = \mathbf{c}^\dagger \hat{H} \mathbf{c}. \tag{2.1.3.1}$$

In the above, $\mathbf{c}^\dagger$ is interpreted as a vector of *creation operators*, that is, it is a vector of M operators $c_i^\dagger$, with $i \in 0, ..., M-1$, that create an electron at lattice site i. If there is already an electron at lattice site i, it will return the number 0. This is known as the *Pauli exlusion principle* (see for example appendix A or [19]). We interpret $\mathbf{c}$ as a vector of *annihilation operators* c_j, with $j \in 0, ..., M-1$, that remove an electron at lattice site j if there is an electron at lattice site j; if this lattice site is not occupied, these operators will return the number 0.

The Hamiltonian $\mathcal{H}$ has the property that $\mathcal{H} = \Sigma_{i,j=0}^{M-1} \alpha_{ij} c_i^\dagger c_j$, for a set of real numbers α_{ij}. This means that the creation operators are always to the left of the annihilation operators (this is called *normal ordering*) and that there is an equal amount of creation and annihilation operators (this makes sure the number of particles is conserved).

The diagonal elements of the Hamiltonian matrix $\hat{H}$ give a certain energy to a lattice site and the i-th diagonal element will be referred to as the *local chemical potential* at site i. The laws of quantum mechanics require $\hat{H}$ to be Hermitian (i.e. equal to the complex conjugate of its transpose). Since we want energies and occupation numbers to be real, $\hat{H}$ will be a real symmetric matrix. We will call the matrix element α_{ij} at position (i, j) (with $i \neq j$) in the Hamiltonian matrix the *hopping* between lattice site i and lattice site j.

A key property of tight-binding models is the assumption that the hopping between lattice sites more than a few sites apart is zero. This means that electrons will only be influenced by the surroundings and not move more than a few sites per time unit. This will lead to a sparse Hamiltonian matrix, a property useful for numerical methods.[7]

2.1.3.1 Tight-binding and finite element methods

Tight-binding models are not completely unrelated to the nearly free electron model. We now demonstrate, following [13], how applying a finite difference scheme to an effective mass equation leads to a tight-binding model.

For simplicity we consider a one-dimensional lattice with no vector potential. If we also absorb the band energy into the potential, (2.1.2.3) simplifies to solving the eigenvalue problem for the Hamiltonian operator $\hat{H}_{op}$

$$\hat{H}_{op} = -\frac{\hbar^2}{2m_n}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + U(x). \tag{2.1.3.2}$$

If we let $f(x)$ a test function, discretise the coordinate x with lattice spacing a and define $f_j := f(ja)$ and $U_j = U(ja)$ for any integer j, we can use a standard finite difference approximation[8]

$$\hat{H}_{op} f \approx -\frac{\hbar^2}{2m_n}\left(\frac{f_{j+1} - 2f_j + f_{j-1}}{a^2}\right) + U_j f_j.$$

If we assume the length to be finite, this becomes a matrix-vector multiplication, where the Hamiltonian matrix is given by

$$\hat{H}_{mat} = \begin{pmatrix} \dots & -\frac{\hbar^2}{2m_n a^2} & 0 & 0 \\ -\frac{\hbar^2}{2m_n a^2} & U_j + \frac{\hbar^2}{m_n a^2} & -\frac{\hbar^2}{2m_n a^2} & 0 \\ 0 & -\frac{\hbar^2}{2m_n a^2} & U_{j+1} + \frac{\hbar^2}{m_n a^2} & -\frac{\hbar^2}{2m_n a^2} \\ 0 & 0 & -\frac{\hbar^2}{2m_n a^2} & \dots \end{pmatrix}. \tag{2.1.3.3}$$

This Hamiltonian matrix $\hat{H}_{mat}$ can also be interpreted as a tight-binding Hamiltonian matrix with local chemical potentials $U_j + \frac{\hbar^2}{m_n a^2}$ and nearest-neighbour hopping terms $-\frac{\hbar^2}{2m_n a^2}$.

2.1.4 The effects of disorder

Until now we have only concerned ourselves with ideal crystals, with perfect symmetries. However, this complete perfection can not be achieved

[7] See also chapter 4 and appendix C.

[8] See for example [21]

in nature, since the lattice structure will always be distorted by for example thermal fluctuations. These distortions of the crystal structure, either caused by chemical impurities (other atoms or molecules than those the lattice consists of) or by crystal defects (the correct atoms or molecules are not at the place they would be in a perfect crystal) are known as *disorder*.[9]

The effect of disorder on the electrons of a crystal is that it in general hinders movement, by making the crystal less uniform. As long as the disorder is not too large, we can still use the effective mass equation (2.1.2.3), but we should add random fluctuations to the potential $U(\mathbf{x})$. Hence we can still use tight-binding models, but with random fluctuations added to the local chemical potential.[10]

2.1.4.1 The Drude model

The Drude model is a classical model for electron transport which expresses the resistance of a material with chemical potential within the conduction band in terms of the disorder, leading to some form of Ohm's law ($R = \frac{V}{I}$ or $\mathbf{J} = \sigma\mathbf{E}$, where R is resistance, I an electrical current, V the applied electric potential, $\mathbf{J}$ the current density, σ the conductivity and $\mathbf{E}$ the electric field). In this model we assume that electrons move classically (so they respond to the force applied to them by an electric field according to Newton's second law) and that the only other way their momentum changes is by colliding with impurities, after which we assume their momentum to be completely random. Following [13], we derive how these assumptions lead to Ohm's law.

We consider a system in a constant electric field $\mathbf{E}$. The electron momentum will converge to a steady state situation quickly and it is this steady state we study here. In the steady state, the expectation value of the electron momentum $\mathbf{p}$ does not change any more:

$$0 = \left\langle \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \right\rangle = e\mathbf{E} - \left\langle \frac{m\mathbf{v}}{\tau} \right\rangle \tag{2.1.4.1}$$

where e is the charge of the electron, m the (effective) mass, $\mathbf{v}$ the electron velocity and τ the time between collisions with impurities. We now define the *momentum relaxation time* $\tau_m := \langle\tau\rangle$ as the mean time between collisions and the electron drift velocity $\mathbf{v_d} := \langle\mathbf{v}\rangle$. We remark that the electric current density can be expressed in terms of electron drift velocity and electron density n_e through $\mathbf{J} = e\mathbf{v_d}n_e$, after which we arrive at Ohm's law

$$\mathbf{J} = e\mathbf{v_d}n_e = \frac{e^2 n_e \tau_m}{m}\mathbf{E}, \tag{2.1.4.2}$$

[9] If there is so much disorder that no crystal structure can be defined any more, we refer to a solid as being amorphous

[10] This makes the term *local* chemical potential appropriate indeed.

where we can see that the relation between current density and electric field is indeed linear and the conductance is given by $\frac{e^2 n_e \tau_m}{m}$, which goes to 0 if the disorder becomes very large and becomes infinite in pure samples.

2.1.4.2 Localisation

We have discussed a model for conductance in which the electrons are treated classically. However, their movement is in general not obeying classical equations; the electrons are quantum mechanical particles. Quantum mechanical effects will change the way electrons move through the system, leading to deviations from Ohm's law.

If the dimension of the system is sufficiently low or the disorder is large, the electronic states will get localised,[22], [23], that is, their position probability density distribution $|\Psi(\mathbf{x})|^2$ will be exponentially decaying away from a certain central point $\mathbf{x_0}$

$$|\Psi(\mathbf{x})|^2 \leq Ce^{-\frac{2|x-x_0|}{\xi}}, \tag{2.1.4.3}$$

where C is some constant and ξ is known as the *localisation length* [22], [23].

The conductance G is defined as the inverse of the resistance, that is, if I is an electrical current, V an applied voltage and R the electric resistance, then $G = \frac{1}{R} = \frac{I}{V}$. Suppose our system is a hypercube of size L^d (L is the length, d the dimension). We define the generalized dimensionless conductance $g(L)$ [23], [24]

$$g(L) := \frac{G}{e^2/2\hbar} \approx \sigma L^{d-2}, \tag{2.1.4.4}$$

where σ is the conductivity and G the conductance. We remark that the approximation in (2.1.4.4) is valid for large L, since this relation holds for uniform conductors, like the systems discussed in section 2.1.4.1.[11] We can also approximate $g(L)$ in the regime where the disorder becomes very large, so that there is localisation ([22], [23], [24])

$$g(L) \sim e^{-\alpha L}, \tag{2.1.4.5}$$

where α is some positive constant. We will now use the scaling argument of [24] to argue that if we are in 1 or 2 dimensions, $g(L \to \infty) = 0$. We first consider multiple identical systems of size L; they have conductance $g(L)$. If we combine b^d such systems to form a system with sides bL, the conductance should be given by $g(bL)$.

[11] Since for uniform and cubic systems $J = \frac{I}{L^{d-1}}$ and $E = \frac{V}{L}$, this is just Ohm's law.

We remark that we want this to hold for any $b \in \mathbb{R}_+$ and that both g and L are always positive, so their logarithms are well defined. Furthermore, $L \mapsto g(L)$ is a smooth bijection from $\mathbb{R}_+$ to $g(\mathbb{R}_+)$. Hence there exists a smooth monotonic function of one variable $\beta(g)$[24], such that

$$\frac{\mathrm{d}\ln g}{\mathrm{d}(\ln L)} = \beta(g(L)).$$

The asymptotics of $\beta(g)$ are obtained from the different approximations for $g(L)$ (2.1.4.4) and (2.1.4.5)

$$\lim_{g\to\infty} \beta(g) = d - 2, \tag{2.1.4.6}$$

$$\lim_{g\to 0} \beta(g) = \ln(Kg), \tag{2.1.4.7}$$

where K is some constant. We conclude that for systems of dimension 1 or 2, $\beta(g)$ is always negative. Since g is a decreasing function of L, the conclusion is that the conductance always vanishes for large systems, so there can be no real metals in 1 or 2 dimensions. We demonstrate this with numerical examples in appendix B.

2.2 The Landauer-Büttiker formalism for electronic transport

We have described how electrons in general behave in solids. We will now model systems that are more related to systems studied in experiments and which consist of an object to be studied (the central region) and wires (to which we will refer as leads) that are connected to the object. The approach will be based on considering this as a scattering problem; it was first developed by Landauer [25] and expanded by Büttiker [26]. This will only be a short introduction; [13] explains this in much more detail, making links to experiments. A rigorous derivation of the important Landauer formula can be found in [27].

2.2.1 Assumed experimental set-up

We assume that the central region and leads are crystalline solids. The ends of the wires would then be connected to things like power sources and measurement devices. We use a model in which we assume the leads to be much longer than the object (we refer to this property as *semi-infinite* and model it by sending one edge of each lead to $\pm\infty$ respectively) and that they are described by the Schrödinger equation with the discrete Schrödinger operator (2.1.3.3) and constant potential (this property is called *ideal*). We assume the leads and central region to be aligned along one direction which we will call the x-axis. A sketch of such a setup is shown in figure 2.1.

Figure 2.1: A sketch of the assumed setup. All systems are discrete and the leads extend to infinity. A wave enters from the left (black arrow) and we measure the part that is transmitted to infinity through the right lead (red arrow)

The experiment now consists of shooting an electron into the lead at $x \to -\infty$ and measuring whether it comes out of the left lead again at $x \to -\infty$ or at the right lead at $x \to \infty$. There is no time scale in this experiment, but otherwise we would also let the time $t \to -\infty$ at the start of the experiment and $t \to \infty$ at the end of the experiment, which is the standard way of modelling scattering in quantum mechanics, see for example [19], [28].

2.2.2 The transport model

Because we are actually interested in the behaviour of the central region as a part of the system, we will choose the leads to be identical and as simple as possible, such that we can treat them exactly or with only a little computational effort.

2.2.2.1 The central region

The part of interest is the central region, which will be described by some model tight-binding Hamiltonian. The central region can be 1-, 2- or 3-dimensional and have any band structure. We pick our leads to have the same dimension as the insulator and a band structure that is (slightly) wider, that is, for any energy in the band of our central region, we want that energy to also be in the band for the leads. All-in-all, the central region can basically be anything we like, as long as it can be described by a tight-binding model and fits into the scattering formalism.

2.2.2.2 Semi-infinite leads

The leads are used to transport electrons to and from the insulator. They will be chosen to be as simple as possible. As simple as possible means that there dimensionality should be the same as that of the insulator, they should be discrete (to make the coupling straightforward) and should be described by a discretised version of (2.1.2.3), with zero vector potential and a

constant potential U and effective mass m chosen such that the band width encompasses all of the insulator spectrum.

We take the lead lattice spacing constant and at first impose zero Dirichlet boundary conditions on a rectangular domain. If our dimensionality is greater than 1, we pick our boundaries in the directions other than the x-direction to be such that the boundaries of the central region are contained within. One boundary in the x-direction is chosen to be the beginning (or end) of the insulator and for the other we take the limit of $x \to \pm\infty$ by adding extra lattice points and keeping the lattice spacing constant. In one dimension we can solve this exactly (see section 3.1.3) and in two or three dimensions we can straightforwardly generalize from this case.

2.2.3 The transmission function

The transmission function T_{pq} is defined to be the chance that an electron shot in through mode q ends up in mode p. We define the transmission function between leads A and B, $T_{AB}(E)$ to be the sum over all modes $p \in A, q \in B$. We remark that $T_{AB}(E)$ takes values between 0 and the number of modes. It is a function of the energy E and we require that $T_{BA} = T_{AB}$.

The Landauer-Büttiker formalism relates the transmission function between **different** leads A, B to the conductance $G(E)$ of an insulator, through the *Landauer formula* [25], [27]

$$G(E) = \frac{e^2}{h} T_{AB}(E). \tag{2.2.3.1}$$

The Landauer formula only holds for systems at zero temperature (but it can be generalized to systems with non-zero temperature) and we assumed that all electrons in the lead have the same spin (else we should multiply the right-hand side of (2.2.3.1) with a factor 2). For non-zero temperature t, the generalisation of the Landauer formula is straightforward, provided that we still apply an infinitesimal voltage bias[12] between the leads [13]. If we let $f_0(E) = \left(e^{\frac{E-\mu}{kt}} + 1\right)^{-1}$ (with μ the chemical potential and k Boltzmann's constant) the equilibrium Fermi-Dirac distribution we get the conductance

$$G(\mu, t) = \frac{e^2}{h} \int T_{AB}(E) \left(-\frac{\partial f_0(E, \mu, t)}{\partial E} \right) \mathrm{d}E. \tag{2.2.3.2}$$

[12]The assumption of an infinitesimal bias can also be lifted, see [13], [14]; in chapter 3 we discuss how we apply a non-equilibrium theory to the equilibrium systems that we want to study.

It can be seen that if $t \to 0$, the two expressions for the conductance coincide (if we identify the Fermi energy $E = \mu$).

2.2.4 The scattering matrix

Since the leads are quantum mechanical systems, we can decide to focus on the waves that travel through them. When these waves hit the scattering region in the middle, their amplitudes are modified and one mode may partly change into another. If the total number of modes in the leads is M, we can define an M-by-M matrix that relates the current amplitudes of these waves before and after the scattering to another.[13] This matrix is known as the S-matrix.

If the current is to be conserved, the S-matrix S is a unitary matrix (see [13] for a derivation). It is clear that there should exist a simple relation between the transmission function and the S-matrix elements; the transmission function between modes p and q and the S-matrix element S_{pq} are related by $T_{pq} = |S_{pq}|^2$. This means that we can indeed obtain the conductance by considering electron transport as a scattering problem.

[13] The current amplitude characterizes the current carried by a wave and is given by the square root of the wave velocity times the wave amplitude.

Chapter 3

Electron transport theory with Green's functions

We have introduced some theory on the transport of electrons on crystals and described the Landauer-Büttiker formalism to connect this electron transport to physical observables. The question of how to compute the relevant transmission function or scattering matrix remains. Green's functions will be introduced and shown to be useful in computing these quantities. This will be further illustrated by examples.

3.1 Green's functions

We start with introducing Green's functions and compute some of importance to our purposes. In section 3.2.1 we need an expression for the Green's function of the infinite discrete ideal lead and in section 3.2.2 we need the Green's function for the semi-infinite discrete ideal lead. Our discussion will follow that in reference [13].

We are interested in solving the time-independent Schrödinger equation, which is in our case either a second order elliptic partial differential equation or a system of linear equations.

We can solve an equation of the form $\hat{A}\mathbf{x} = \mathbf{b}$, where $\hat{A}$ is a given invertible matrix, $\mathbf{b}$ a given vector and $\mathbf{x}$ an unknown vector by inverting $\hat{A}$: $\mathbf{x} = \hat{A}^{-1}\mathbf{b}$. The method of Green's functions consists of proceeding analogously in trying to find a solution to a differential equation; a Green's function is a distribution that is a right inverse of a differential operator.[1] The Green's functions we need are not unique, because we will still have to

[1] We will actually also refer to the right inverse of finite difference operators and sometimes even the inverses of matrices as Green's functions, if these finite difference operators and matrices arise from tight-binding models.

enforce boundary conditions.

Suppose we have a system described by an effective mass equation (2.1.2.3), to which we add an excitation $f(\mathbf{x})$ (i.e. an incoming wave from the leads). We will then have an equation of the form

$$\left(\hat{H}_{op} + E_n - E\right)\Psi = f. \tag{3.1.0.1}$$

Loosely speaking, we will then take $G = -\left(\hat{H}_{op} + E_n - E\right)^{-1}$ and $\Psi = -Gf$.[2] To make this more precise, we need to introduce some theoretical concepts regarding partial differential equations. The introduction here will be brief, but much more background information can be found in for example [29], [30] and [31].

3.1.1 What are Green's functions?

When we look at the operator $\hat{H}_{op}$, a natural question to ask are what its domain and co-domain are. It is clear that the Hamiltonian sends a function to another function, so we will now define some important function spaces known as *L^p-spaces* and their generalizations, *Sobolev spaces*.

Definition 1. *Let $1 \leq p < \infty$, $n \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^n$ open. Let $f : \Omega \to \mathbb{R}$ be a function. The $L^p(\Omega)$-*norm *is defined by $||f||_{L^p(\Omega)} := \left(\int_\Omega |f(x)|^{\frac{1}{p}}\,\mathrm{d}x\right)^p$.*

We remark that all integrals are to be understood as Lebesgue integrals and that with "function" we may also mean "the equivalence class of functions that are the same almost everywhere", that is, functions that only differ from each other on a set of measure 0. An introduction to measure theory can for example be found in [32].

Definition 2. *Let $1 \leq p < \infty$, $n \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^n$ open. The space of functions with finite $L^p(\Omega)$-norm, $\left(L^p(\Omega), ||.||_{L^p(\Omega)}\right)$, is called $L^p(\Omega)$.*

Proposition 1. *$L^p(\Omega)$ is a Banach space. $L^2(\Omega)$ is a Hilbert space.*

Proof. The proof can be found in many books, for example in [29]. □

Sobolev spaces are generalizations of L^p-spaces, in the sense that a function is an element of a Sobolev space if a certain derivative of it is in an L^p-space, or if it is a certain derivative of a function in an L^p-space. Before we can properly introduce Sobolev spaces, we should introduce the concept of weak derivatives.

[2] It is conventional to choose the signs like this.

Definition 3. *Let $n \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^n$ open. Let $f : \Omega \to \mathbb{R}$ be a function. We say that $g : \Omega \to \mathbb{R}$ is a* weak derivative *of f, if $\int_\Omega f\phi' = -\int_\Omega g\phi \qquad \forall \phi \in C_0^\infty(\Omega)$, where $C_0^\infty(\Omega)$ is the space of infinitely differentiable and compactly supported functions on Ω and ϕ' is the derivative of the* test function *ϕ.*

Proposition 2. *For any differentiable function $u : \Omega \to \mathbb{R}$ the strong derivative u'_s coincides with the weak derivative u'_w almost everywhere.*

Proof. Let ϕ be a test function. Then $\int_\Omega (u'_s - u'_w)\,\phi = \int_\Omega u'_s\phi - \int_\Omega u'_w\phi = \int_\Omega u\phi' - \int_\Omega u\phi' = 0$, where we first used the linearity of the integral and then the definition of the weak derivative and partial integration. We conclude that $u'_s = u'_w$ almost everywhere by the fundamental lemma of the calculus of variations. □

The above proposition shows that weak derivatives are indeed a generalisation of ordinary (strong) derivatives. In the same way, Sobolev spaces are a generalisation of C^k-spaces

Definition 4. *The* Sobolev space *$W^{k,p}(\Omega)$, where $k \in \mathbb{N}$ and $p \in [1, \infty]$, is defined as the space of functions $f \in L^p(\Omega)$ for which the k-th weak derivative is also in $L^p(\Omega)$. The* Sobolev norm *is defined as $||f||_{W^{k,p}(\Omega)} := \left(\sum_{j=0}^k ||D^j f||^p_{L^p(\Omega)}\right)^{\frac{1}{p}}$.*

Proposition 3. *Sobolev spaces with Sobolev norm defined above are Banach spaces. If $p = 2$ they are Hilbert spaces.*

Proof. The proof can be found in many books, for example in [29]. □

We can now see that the differential operator H_{op} will send an element of $W^{k+2,p}(\Omega)$ to $W^{k,p}(\Omega)$, for all $k \geq 0$. Sobolev spaces satisfy inequalities known as Sobolev inequalities. In particular, for $p > N \in \mathbb{N}_+$ the following holds

Theorem 2 (Morey)**.** *Let $p > N \in \mathbb{N}_+$. Then $W^{1,p}(\mathbb{R}^N) \subset L^\infty(\mathbb{R}^N)$ with continuous injection. For all $u \in W^{1,p}(\mathbb{R}^N)$, almost all $x, y \in \mathbb{R}^N$, $\alpha = 1 - \frac{N}{p}$ and C a constant depending on p and N*

$$|u(x) - u(y)| \leq C|x - y|^\alpha ||\nabla u||_p.$$

This implies that all such functions u are equal to a continuous function almost everywhere, so we can represent all elements of $W^{1,p}(\mathbb{R}^N)$ by a continuous representative.

Proof. The proof can be found in many books, for example in [29]. □

We can now introduce the concept of distributions and in particular the Dirac δ-distribution. We choose $p = 2$ and represent every element of $W^{1,2}(\mathbb{R}^N)$ by its continuous representative.

Definition 5. *Let $C_0^\infty(\Omega)$ be the space of test functions. A* distribution *is a linear functional on $C_0^\infty(\Omega)$. The* Dirac δ-distribution *is defined by $< \delta, \phi >_{L^2(\Omega)} := \phi(0)$, for $\phi \in C_0^\infty(\Omega)$.*

Proposition 4. *The Dirac δ-distribution is an element of the dual space of $(W^{1,2}(\mathbb{R}^N), ||.||_{L^2(\mathbb{R}^N)})$.*

Proof. By Moreys's theorem we can represent all elements $f(x) \in (W^{1,2}(\mathbb{R}^N), ||.||_{L^2(\mathbb{R}^N)})$ to be continuous and bounded. Then $< \delta, f >_{L^2(\mathbb{R}^N)} = \int_{\mathbb{R}}^N) f(x)\delta(x)\,\mathrm{d}x = f(0)$ defines a bounded linear functional, so the δ-distribution is in the dual space. □

We remark that we have extended the δ-distribution to a functional on $W^{1,2}(\mathbb{R}^N)$ in the above, which can be done unambiguously. We also remark that we use the L^2-norm instead of the $W^{1,2}$-norm and do not identify $W^{1,2}$ with its dual space.

The Dirac δ-distribution plays the role of an identity in the sense that the retarded Green's function is a solution of

$$\left(\epsilon - \hat{H}_{op}(x) + i\eta\right) G(x, x') = \delta(x - x'), \tag{3.1.1.1}$$

where $x, x' \in \mathbb{R}^N$ are coordinates and η is a positive infinitesimal that will make sure our solution is unique[3]. We have also fixed the band index n and defined $\epsilon = E - E_n$. We have now made our loose definition of the Green's function precise: the Green's function is defined by (3.1.1.1) and the solution of (3.1.0.1) is given by $\Psi(x) = -\int_{\mathbb{R}^N} G(x, x')f(x')\,\mathrm{d}x'$.

3.1.2 Green's functions for finite difference operators

We have defined the Green's function for differential operators. However, we will be most interested in using Green's functions to solve finite difference equations and matrix equations. We now relate these different notions of Green's functions and we find a way to compute these Green's function using the *spectral representation*.

As explained in section 2.1.3.1, finite difference operators are both related to differential operators and to matrices. This suggests that their right inverse

[3] (and satisfies certain boundary conditions if we use it for evaluating the time-*dependent* Schröedinger equation)

should be something that has some properties of Green's functions and also some properties of inverse matrices. For a matrix, the inverse would be a matrix dependent on the real parameter ϵ. In the previous section we have found that the Green's function for a differential operator depends on the real numbers x, x', ϵ.

We can connect these two by realizing that every matrix element has two indices corresponding to two lattice sites. This suggests that the Green's function for a finite difference operator should also depend on the lattice sites and energy, so it will be a function $G : I \times I \times \mathbb{R}$, where I is an (countably infinite) index set. If we would take I to be finite, we would recover a matrix, whereas taking I to be uncountable would give us the case of a Green's function depending on continuous coordinates, i.e. a Green's function for a differential operator.

It is important to note that the Green's function of a finite difference operator is a right inverse; in general it will not commute with the operator to also be a left inverse. This is the same as for differential operators, but different from the way matrices work. We can use a discrete version of (3.1.1.1) to define the retarded Green's function for finite difference operators: we replace x, x' by indices $j, l \in I$ and replace the Dirac δ-distribution by a Kronecker δ-function δ_{jl}, which is 1 if $j = l$ and 0 otherwise.

In order to compute the retarded Green's function we will derive the *spectral representation* of the Green's function. Suppose we have solved the eigenvalue problem for the Hamiltonian operator H_{op}. We then have a complete orthonormal basis $\{\Psi_n\}_{n\in\mathbb{N}}$ of eigenfunctions of the Hamiltonian operator H_{op} with corresponding eigenvalues $\{E_n\}$. We define projectors Π_n which project any state on the corresponding basis element (in Dirac notation these projectors are defined as $\Pi_n := |\Psi_n\rangle \langle\Psi_n|$, see appendix A or [19]). Any state Ψ can now be decomposed into basis states: $|\Psi\rangle = \sum_{n\in\mathbb{N}} \Pi_n |\Psi\rangle$.

In particular, we can decompose the Green's function $|G\rangle := G(\epsilon, j; l)$ de-

fined by the discrete version of (3.1.1.1) into basis states. Then

$$\begin{aligned}\delta_{jl} = \left(\epsilon - \hat{H}_{op}(j) + i\eta\right) G(\epsilon, j; l) &= \left(\epsilon - \hat{H}_{op} + i\eta\right) |G\rangle \qquad (3.1.2.1)\\ &= \left(\epsilon - \hat{H}_{op} + i\eta\right) \sum_{n\in\mathbb{N}} \Pi_n |G\rangle \\ &= \sum_{n\in\mathbb{N}} \left(\epsilon - \hat{H}_{op} + i\eta\right) |\Psi_n\rangle \langle\Psi_n, G\rangle \\ &= \sum_{n\in\mathbb{N}} (\epsilon - E_n + i\eta) |\Psi_n\rangle \langle\Psi_n, G\rangle \\ &= \sum_{n\in\mathbb{N}} (\epsilon - E_n + i\eta) \Pi_n |G\rangle . \qquad (3.1.2.2)\end{aligned}$$

It is now clear that we can write $G(\epsilon, j; l)$ as

$$\begin{aligned}G(\epsilon, j; l) &= \sum_{n\in\mathbb{N}} \frac{\Pi_n(j, l)}{\epsilon - E_n + i\eta} \\ &= \sum_{n\in\mathbb{N}} \frac{\Psi_{j,n}\Psi^*_{l,n}}{\epsilon - E_n + i\eta}. \qquad (3.1.2.3)\end{aligned}$$

The spectral representation (3.1.2.3) is a useful tool for the computation of Green's functions for which the eigenvalue problem is (easily) solvable.

3.1.3 The Green's function of an ideal lead

We compute the Green's functions for one-dimensional discrete ideal leads for both infinite and semi-infinite leads. The discrete lead is described by (2.1.3.3), with potential energy U_i set to 0.

We start by considering the discrete lead with Dirichlet boundary conditions at the boundary points 0 and L. The equations become as follows

$$\begin{aligned}-\frac{\hbar^2}{2ma^2}\Psi_{j-1} + \frac{\hbar^2}{ma^2}\Psi_j - \frac{\hbar^2}{2ma^2}\Psi_{j+1} = E\Psi_j, \qquad (3.1.3.1)\\ \Psi_0 = 0, \quad \Psi_N = 0,\end{aligned}$$

where $\Psi_j = \Psi(ja)$ and $N = L/a$. Lattice points are at ja for $j \in 0, 1, ..., N$ and will also be denoted by $x_j := ja$. It can be useful to view the system as a tight-binding model instead of a discretisation of a continuous model (see section 2.1.3.1), so we also introduce the parameters $\mu = \frac{\hbar^2}{ma^2}$ and $t = -\frac{\hbar^2}{2ma^2}$.

We note that the eigenfunctions of a corresponding continuous ideal lead

are sinusoids, so $\Psi_{\kappa_n}(x) \sim \sin(\kappa_n x)$ (where $\kappa_n = \frac{\pi n}{L}$). For the discrete lead we make the same ansatz for the eigenfunctions and plug them into (3.1.3.1)

$$\begin{aligned} E\sin(\kappa_n x_j) &= -\frac{\hbar^2}{2ma^2}\sin(\kappa_n x_{j-1}) + \frac{\hbar^2}{ma^2}\sin(\kappa_n x_j) - \frac{\hbar^2}{2ma^2}\sin(\kappa_n x_{j+1}) \\ &= -\frac{\hbar^2}{2ma^2}\sin(\kappa_n (x_j - a)) + \frac{\hbar^2}{ma^2}\sin(\kappa_n x_j) - \frac{\hbar^2}{2ma^2}\sin(\kappa_n (x_j + a)) \\ &= \frac{\hbar^2}{ma^2}(1 - \cos(\kappa_n a))\sin(\kappa_n x_j). \end{aligned} \tag{3.1.3.2}$$

This shows that the eigenfunctions are indeed given by $\Psi_{j,n} = \sqrt{2/L}\sin(\kappa_n x_j)$, with dispersion relation $E(\kappa_n) - \mu = 2t\cos(\kappa_n a)$. The velocity of the wave packets is given by

$$v_n = \frac{1}{\hbar}\frac{\partial E}{\partial \kappa_n} = \frac{\hbar}{ma}\sin(\kappa_n a). \tag{3.1.3.3}$$

We can now compute the Green's function using the spectral representation (3.1.2.3)

$$G^R(\epsilon, j; l) = \sum_{n=1}^{N} \frac{\Psi_{j,n}\Psi^*_{l,n}}{\epsilon - E(\kappa_n) + i\eta} \tag{3.1.3.4}$$

$$\begin{aligned} &= \frac{2}{L}\sum_{n=1}^{N} \frac{\sin(\kappa_n x_j)\sin(\kappa_n x_l)}{\epsilon - \frac{\hbar^2}{ma^2}(1 - \cos(\kappa_n a)) + i\eta} \\ &= \frac{2}{L}\sum_{n=1}^{N} \left\{ \frac{\frac{e^{i\kappa_n x_j} - e^{-i\kappa_n x_j}}{2i}\frac{e^{i\kappa_n x_l} - e^{-i\kappa_n x_l}}{2i}}{\epsilon - \frac{\hbar^2}{ma^2}(1 - \frac{e^{i\kappa_n a} + e^{-i\kappa_n a}}{2}) + i\eta} \right\}. \end{aligned} \tag{3.1.3.5}$$

We introduce the variable $z_n = e^{i\kappa_n a}$ and take the limit $L \to \infty$, which enables us to write the retarded Green's function as an integral over the unit circle in the complex plane

$$G^R(\epsilon, j; l) = \frac{-i}{4\pi a}\oint_{|z|=1} \frac{\left(z^j - z^{-j}\right)\left(z^l - z^{-l}\right)}{\epsilon - \frac{\hbar^2}{ma^2}\left(1 - \frac{1}{2}\left(z + \frac{1}{z}\right)\right) + i\eta}\frac{\mathrm{d}z}{z}. \tag{3.1.3.6}$$

We can compute this integral by means of Cauchy's residue theorem (see for example [33]). In order to do this it is helpful to split the powers of z in the numerator and identify where the poles are for general powers $p \in \mathbb{Z}$ of z, so we should solve

$$z^p\left(z^2 + \left(\frac{2ma^2\epsilon}{\hbar^2} - 2 + i\delta\right)z + 1\right) = 0, \tag{3.1.3.7}$$

where the infinitesimal $\delta = \frac{2ma^2}{\hbar^2}\eta$. We define $\bar{\epsilon} := \frac{2ma^2\epsilon}{\hbar^2}$. For $p \geq 1$ there is a pole of order p at 0, a pole of order 1 somewhere within the unit circle

and another pole of order 1 outside of the unit circle. For $p \leq 0$ there is one pole of order 1 within and one pole of order one outside of the unit circle, but there are no poles at 0. The poles of order 1 are given by

$$z_{\pm} = -\frac{\bar{\epsilon}}{2} + 1 - \frac{i\delta}{2} \pm \frac{1}{2}\sqrt{\bar{\epsilon}^2 - 4\bar{\epsilon} + 2i(\bar{\epsilon} - 2)\delta}. \tag{3.1.3.8}$$

We remark that $z_+ = \frac{1}{z_-}$, so there will always be one pole within and one pole outside of the unit circle in the complex plane.[4] We can check numerically that z_+ is inside the unit circle if $\bar{\epsilon} \in [2, \infty[$ and that z_- is in the unit circle when $\bar{\epsilon}$ is not in this interval. Another property of $z_{\pm}$ is that it only has a non-zero imaginary part if $\bar{\epsilon} \in [0, 4]$, that is, if the energy is within the band width.

We must in addition make the important remark that the square root has a branch cut along the negative real axis. This means that the $2i(\bar{\epsilon} - 2)\delta$ will give an extra minus sign in front of the square root for $\bar{\epsilon} \in [0, 2]$. In practice this means that when we take $\delta \to 0$ we should select the solution with z_+ for the interval $[0, \infty]$, in this way accounting for the extra minus sign automatically.

This now implies that the computation of (3.1.3.6) can be performed by considering

$$\oint_{|z|=1} \frac{\mathrm{d}z}{z^p(z - z_+)(z - z_-)} = \begin{cases} \frac{2\pi i}{z_+^p(z_+ - z_-)}, & \text{if} \quad p \leq 0 \text{ and } |z_+| < 1, \\ \frac{2\pi i}{z_-^p(z_- - z_+)}, & \text{if} \quad p \leq 0 \text{ and } |z_-| < 1, \\ \frac{2\pi i}{z_+^p(z_+ - z_-)} + \frac{\mathrm{d}^{p-1}}{\mathrm{d}z^{p-1}} \frac{2\pi i}{(z - z_+)(z - z_-)}\Big|_{z=0}, & \text{if} \quad p \geq 1 \text{ and } |z_+| < 1, \\ \frac{2\pi i}{z_-^p(z_- - z_+)} + \frac{\mathrm{d}^{p-1}}{\mathrm{d}z^{p-1}} \frac{2\pi i}{(z - z_+)(z - z_-)}\Big|_{z=0}, & \text{if} \quad p \geq 1 \text{ and } |z_-| < 1. \end{cases} \tag{3.1.3.9}$$

We rewrite (3.1.3.6) in the form of (3.1.3.9)

$$G^R(\epsilon, j; l) = -\frac{ima}{2\pi\hbar^2} \left(\oint_{|z|=1} \frac{\mathrm{d}z}{z^{-(j+l)}(z - z_+)(z - z_-)} - \oint_{|z|=1} \frac{\mathrm{d}z}{z^{-(j-l)}(z - z_+)(z - z_-)} - \oint_{|z|=1} \frac{\mathrm{d}z}{z^{(j-l)}(z - z_+)(z - z_-)} + \oint_{|z|=1} \frac{\mathrm{d}z}{z^{(j+l)}(z - z_+)(z - z_-)} \right). \tag{3.1.3.10}$$

In principle we have now computed the semi-infinite lead retarded Green's

[4] The infinitesimal δ will guarantee that there are no poles *on* the unit circle.

function for any two lattice points. To provide some further insight and for later reference, we will now set $j = l = 1$ and work out the derivatives in (3.1.3.9)

$$G^R(\epsilon, 1; 1) = \frac{ma}{\hbar^2}\left(\pm\frac{z_\pm^2}{(z_+ - z_-)} \mp \frac{2}{(z_+ - z_-)} \pm \frac{z_\pm^{-2}}{(z_+ - z_-)} + (z_+ + z_-)\right), \tag{3.1.3.11}$$

where the $\pm$-sign is determined by the pole $z_\pm$ that is within the unit circle.

We can now define the momentum $\kappa_L(\epsilon) \in \mathbb{C}$ by choosing $z_+(\epsilon) = e^{i\kappa_L a}$. Then $2\cos(\kappa_L a) = z_+ + z_- = (2 - \bar{\epsilon})$ and we re-obtain our dispersion relation and lead velocity

$$\epsilon - \mu = 2t\cos(\kappa_L a), \tag{3.1.3.12}$$

$$v_L = \frac{1}{\hbar}\frac{\partial\epsilon}{\partial\kappa_L} = \frac{\hbar}{ma}\sin(\kappa_L a). \tag{3.1.3.13}$$

We can also rewrite our Green's function in terms of κ_L and as follows

$$G^R(\epsilon, 1; 1) = \frac{-1}{ta}e^{\pm i\kappa_L a}, \tag{3.1.3.14}$$

where we remark that both κ_L and v_L are in general not purely real. We have now computed the right lead Green's function, but the left lead Green's function is of course precisely the same.

The computation of the infinite lead Green's function is analogous to the computation of the semi-infinite lead Green's function. In this case the eigenfunctions are given by $\Psi_{j,n} = \sqrt{1/L}e^{i\kappa_n x_j}$, where n now runs from $-N$ to N. The dispersion relation remains the same. We compute the Green's function in the same way, remarking that we now go round the contour twice. The integral analogous to (3.1.3.6) is

$$G^R_{Inf}(\epsilon, j; l) = \frac{i}{\pi a}\oint_{|z|=1} \frac{z^{j-l}}{\epsilon - \frac{\hbar^2}{ma^2}\left(1 - \frac{1}{2}\left(z + \frac{1}{z}\right)\right) + i\eta}\frac{\mathrm{d}z}{z}. \tag{3.1.3.15}$$

If we now consider the case where $j = l$ we obtain

$$G^R_{Inf}(\epsilon, j; j) = \frac{2ma}{\hbar^2}\frac{1}{z_+ - z_-} = \frac{-ima}{\hbar^2}\frac{1}{\sin(\kappa_L a)} = -\frac{i}{\hbar v_L}. \tag{3.1.3.16}$$

3.2 The non-equilibrium Green's function formalism

The non-equilibrium Green's function formalism is a way of describing quantum transport. It was originally developed by Kadanoff and Baym [11]

and by Keldysh [10]. Another name for it is Keldysh formalism. For non-interacting systems in equilibrium this formalism is equivalent to the Landauer-Büttiker formalism described previously (this will be shown in the part on Fisher-Lee relations).

3.2.1 Fisher-Lee relations

Fisher-Lee relations [12] describe the scattering matrix in terms of Green's functions and in this way connect the non-equilibrium Green's function formalism to the Landauer-Büttiker formalism.

A system as in the section on the Landauer-Büttiker formalism can be described as an ideal lead that is modified such that all wave functions are multiplied by the wave amplitude scattering matrix upon passing through the origin. The (part of the) lead to the left of the origin will be referred to as lead $-$ and the part to the right of the origin as lead $+$.

If we consider the retarded Green's function for a one-dimensional infinite lead derived above in a lattice point 0_- with coordinate smaller then but infinitesimally close to 0, the left-moving amplitude is not modified, while the right-moving amplitude is scattered and the exponential becomes 1. Hence [13],

$$G^R_{\pm}(0_\pm; 0_-) = \delta_{\pm -} A^- + S^w_{\pm -} A^+, \tag{3.2.1.1}$$

where $\delta_{\pm -}$ is the Kronecker delta function and $S^w_{\pm -}$ the wave scattering matrix element corresponding to the scattering by the system. Using (3.1.3.16), $A^+ = A^- = -\frac{i}{\hbar v_-}$ and $S^w_{\pm -} = \sqrt{\frac{v_-}{v_\pm}} S_{\pm -}$ and taking into account when $\delta_{\pm -}$ is 0 we arrive at the Fisher-Lee relation

$$S_{\pm -} = -\delta_{\pm -} + i\hbar\sqrt{v_- v_\pm} G^R_{\pm}(0_\pm; 0_-). \tag{3.2.1.2}$$

In the above Fisher-Lee relation the velocities are given by (3.1.3.12) and (3.1.3.13).

In case of multi-moded leads the following Fisher-Lee relation can be derived [13]

$$S_{nm} = -\delta_{nm} + i\hbar\sqrt{v_n v_m} \iint \chi_n(y_\pm) G^R_{\pm}(0_\pm, y_\pm; 0_-, y_-) \chi_m(y_-)\, \mathrm{d}y_\pm\, \mathrm{d}y_-, \tag{3.2.1.3}$$

where the indices n, m number the channels (n in lead $\pm$ and m in lead $-$), transverse coordinates are labelled with y and transverse wave-functions with χ.

3.2.2 Lead self-energy

We have derived the Green's function for an isolated semi-infinite discrete lead and we described how to compute the scattering matrix from the systems Green's function. We now connect the leads to the insulator and use the expressions for the lead Green's functions to simplify the problem of computing the system Green's function. This will also enable us to derive a more compact expression for the transmission function (and hence the conductance) between two leads.

3.2.2.1 The leads as boundary conditions on the system

Our system consists of an insulator connected on both sides to a semi-infinite lead. The system is described by a tight-binding operator. We can distinguish five different components in the system Hamiltonian, namely the left and right lead Hamiltonians $\hat{H}_l$ and $\hat{H}_r$ (tight binding operators), the insulator Hamiltonian H_i (a matrix, since the insulator is finite) and the respective couplings τ_l, τ_r of the leads to the insulator (of finite rank, since only a finite number of sites is coupled).

Since inverting matrices is in general easier than inverting operators and we have already inverted the lead operators, we would like to use the lead Green's functions to reduce the problem to a matrix inversion problem. This means that we would like to alter the Hamiltonian matrix in such a way that the leads and lead-system couplings are included; if we were to consider the corresponding differential equation problem, this would amount to using special "open" boundary conditions.[5]

We write the system Hamiltonian as a block matrix.[6]

$$\hat{H}_{system} = \begin{pmatrix} \hat{H}_l & \tau_l & 0 \\ \tau_l^\dagger & \hat{H}_i & \tau_r \\ 0 & \tau_r^\dagger & \hat{H}_r \end{pmatrix}. \tag{3.2.2.1}$$

The system retarded Green's function G_{system} is the right inverse of $(E + i\eta)\hat{\mathbf{1}} - \hat{H}_{system}$, where $\hat{\mathbf{1}}$ is the identity operator. G_{system} can also be written in the same form as the Hamiltonian

$$G_{system} = \begin{pmatrix} G_l & G_{li} & G_{lr} \\ G_{il} & G_i & G_{ir} \\ G_{rl} & G_{ri} & G_r \end{pmatrix}.$$

[5] As pointed out in [13], writing the insulator, coupling and a finite part of the lead as a matrix and imposing Dirichlet conditions would correspond to a closed system, where no electrons can get out of. Writing down a finite matrix and making it "infinite" by applying periodic boundary conditions would yield an infinite chain of finite leads coupled to insulators, which is also clearly not what we want.

[6] Note that the lead (and system) Hamiltonians are now infinite-dimensional matrices

If we multiply G_{system} by $\hat{H}_{system}$ from the left we would get the following equations (where $\hat{\mathbf{1}}$ stands for an appropriate identity matrix or operator and $\hat{\mathbf{0}}$ for an appropriate zero matrix or operator)

$$\left((E+i\eta)\hat{\mathbf{1}} - \hat{H}_l\right) G_{li} + \tau_l G_i = \hat{\mathbf{0}}, \tag{3.2.2.2}$$

$$\left((E+i\eta)\hat{\mathbf{1}} - \hat{H}_r\right) G_{ri} + \tau_r^\dagger G_i = \hat{\mathbf{0}}, \tag{3.2.2.3}$$

$$\tau_l^\dagger G_{li} + \left((E+i\eta)\hat{\mathbf{1}} - \hat{H}_i\right) G_i + \tau_r G_{ri} = \hat{\mathbf{1}}. \tag{3.2.2.4}$$

It is clear that (3.2.2.2) and (3.2.2.3) can be solved in terms of the semi-infinite lead Green function. If we then substitute these solutions into (3.2.2.4), we obtain

$$-\tau_l^\dagger \mathcal{G}^L \tau_l G_i + \left((E+i\eta)\hat{\mathbf{1}} - \hat{H}_i\right) G_i - \tau_r \mathcal{G}^R \tau_r^\dagger G_i = \hat{\mathbf{1}}, \tag{3.2.2.5}$$

where $\mathcal{G}^{L(R)}$ is the Green's function of the isolated semi-infinite left (right) lead. These identical quantities have already been computed (see (3.1.3.10) and (3.1.3.14)).

We note that (3.2.2.5) is easier to solve then (3.2.2.2)-(3.2.2.4), because (3.2.2.5) is a matrix-equation for a matrix of dimensions equal to the dimensions of the Hamiltonian matrix; because we have solved the problem of the isolated semi-infinite lead and the coupling between the leads and insulator is only on the area near the insulator boundary, we have reduced the effect of the leads to that of a boundary condition on the insulator. Therefore our focus will be on (efficiently) solving (3.2.2.5) and then applying (3.2.1.2) or (3.2.1.3) to obtain the scattering matrix and hence the conductance.

3.2.2.2 Definition of the lead self-energy and lead velocity operator

In (3.2.2.5) we can combine all the three terms into one such that the insulator Green's function becomes the inverse of a modified matrix

$$\left((E+i\eta)\hat{\mathbf{1}} - \hat{H}_i - \Sigma\right) G_i = \hat{\mathbf{1}}, \tag{3.2.2.6}$$

where we have defined the *retarded lead self-energy* $\Sigma := \Sigma_l + \Sigma_r := \tau_l^\dagger \mathcal{G}^L \tau_l + \tau_r \mathcal{G}^R \tau_r^\dagger$. We can compute the retarded lead self-energy from the analytic lead Green's functions and the couplings and subtract it from $\left((E+i\eta)\hat{\mathbf{1}} - \hat{H}_i\right)$, after which we need to invert the resulting matrix. We note that the retarded lead self-energy has in general a non-zero imaginary component (for instance if there are propagating modes [13], [14]), which means that we can leave the infinitesimal η out of our computations.

Example 1 (Computation of the retarded lead self-energy)**.** *Suppose we have a 1D system of length N for which the only connections between leads and insulator are hoppings with strength t_c between the last site of the left lead and the first site of the insulator and between the last site of the insulator and the first site of the right lead. Then the self-energy matrix only has non-zero elements in the upper-left and lower right corner. These elements are equal and given by $\Sigma_{11}(E) = t_c^2 \mathcal{G}^L(E,1;1)$, where $\mathcal{G}^L(E,1;1)$ is given by (3.1.3.14). We define $\bar{\epsilon} := \frac{2ma^2\epsilon}{\hbar^2}$ and $t = \frac{-\hbar^2}{2ma^2}$ and plot $\frac{at\Sigma_{11}(\bar{\epsilon})}{t_c^2}$ in figure 3.1. The real part is plotted in blue and the imaginary part is plotted in red.*

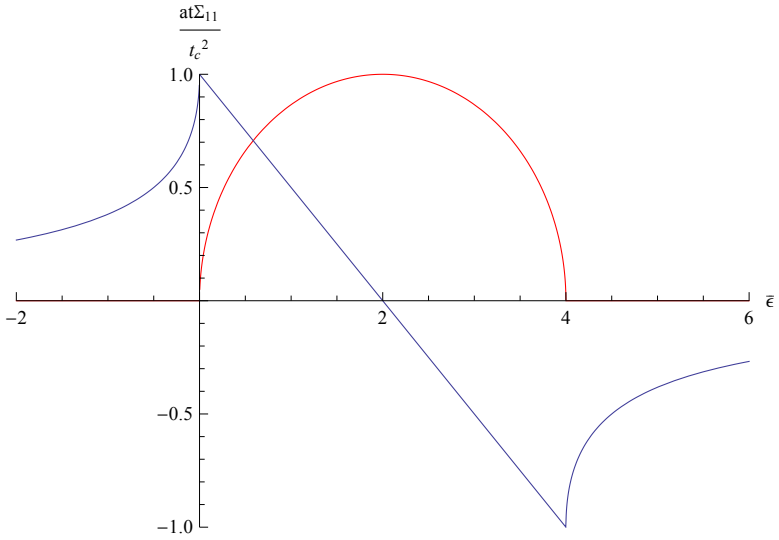

Figure 3.1: A plot of $\frac{at\Sigma_{11}(\bar{\epsilon})}{t_c^2}$ as a function of $\bar{\epsilon}$. Also confer [34]

We can also define the *advanced lead self-energy* Σ^A analogously by taking the equations for the advanced Green's functions corresponding to (3.2.2.2)-(3.2.2.6) instead. We will then get $\Sigma^A := \tau_l \mathcal{G}^{LA} \tau_l^\dagger + \tau_r^\dagger \mathcal{G}^{RA} \tau_r$, where $\mathcal{G}^{L(R)A}$ is the advanced Green's function of the isolated semi-infinite left (right) lead. If $\left(\mathcal{G}^{L(R)A}\right)^\dagger = \mathcal{G}^{L(R)}$, then $\Sigma^A = \Sigma^\dagger$.

With the advanced and retarded lead self-energy we can define the *lead velocity operator* or *level-width function* $\Gamma_{l(r)} := i(\Sigma_{l(r)} - \Sigma^A_{l(r)})$, which is the velocity or current operator within the left (right) lead [14], [34].

3.2.2.3 A compact expression for the conductance

With the help of the lead velocity operator we can find a compact expression for the transmission function T_{lr} between left and right lead. It is given by

$$T_{lr} = \mathrm{Tr}\left[\Gamma_l G^R \Gamma_r G^A\right]. \tag{3.2.2.7}$$

In [13] (3.2.2.7) is derived for multi-dimensional leads that are coupled only at the edge, with coupling constant $t = \frac{\hbar^2}{2ma^2}$, where m is the effective mass inside the lead. In [14] it is shown that this formula also holds for very general systems. Here we verify the formula for systems discussed in example 1 above.

We sketch the argument given by [13]. We note that the matrices $\Gamma_{l(r)}$ only have 1 non-zero element. This element is identical and given by $\Gamma_l^{11} = \Gamma_r^{NN} = -\hbar v_L$. The fact that these matrices only have one non-zero element means that we can scale out this element and perform the matrix multiplication easily: the product $\Gamma_l G^R$ will be (proportional to) a matrix with first row equal to G^R and the rest zeros. Further multiplication by Γ_r will yield a matrix whose only non-zero element is in the upper right corner and proportional to G^R_{1N}. Multiplication by G^A will pick out the last row of G^A after which taking the trace gives

$$\begin{aligned} T_{lr} &= \hbar^2 v_L^2 G^R_{1N} G^A_{N1} \\ &= \hbar^2 v_L^2 G^R_{1N} G^{R*}_{1N} \\ &= |S_{lr}|^2, \end{aligned} \tag{3.2.2.8}$$

where we used the fact that our Green's functions are actually inverse matrices, $\Sigma^A = \Sigma^\dagger$ and the previously derived Fisher-Lee relation (3.2.1.2).

It is important to note that the matrices $\Gamma_{l(r)}$ are very sparse and only contain elements corresponding to connecting the left (right) edge of the insulator to the leads. This means that our matrix products in (3.2.2.7) will also yield very sparse matrices and the traces will only consists of sums of very few elements of the Green's function. An efficient way of computing the transmission function would then be to decide beforehand which of the elements we need and what combination they will form, thereby saving us the effort of the big matrix multiplications and of taking the trace and also enabling us to think of ways to not compute the entire Green's functions, but only the elements we need. These more efficient ways of computing Green's functions will be first addressed in chapter 4; we will now derive a more general formula expressing (3.2.2.7) into (generally) smaller parts of the Green's functions.

Suppose we want to compute Σ_l for given coupling τ_l and isolated (retarded) lead Green's function $\mathcal{G}^L$. Then $\Sigma_l = \tau_l^\dagger \mathcal{G}^L \tau_l$; this self energy is defined for all combinations of sites (i, j) in the system. We denote the set of sites within the lead for which the coupling to system site i is non-zero by $\{p_i\}$ and remark that $\Sigma_l(i, j) = 0$ if either $\{p_i\}$ or $\{p_j\}$ is empty. It follows from the definition of τ_l that it has a finite number of columns and an infinite number of rows. Following the convention that the first index in a matrix denotes the row number and the second the column number, we can straightforwardly write down an expression for the non-trivial elements of the left self energy matrix

$$\Sigma_l(i, j) = \sum_{k \in \{p_i\}} \sum_{n \in \{p_j\}} \tau_l^\dagger(i, k) \mathcal{G}^L(k, n) \tau_l(n, j).$$

For the right self energy matrix the non-trivial elements are given by a similar formula

$$\Sigma_r(i, j) = \sum_{k \in \{p_i\}} \sum_{n \in \{p_j\}} \tau_r(i, k) \mathcal{G}^R(k, n) \tau_r^\dagger(n, j).$$

We have derived that $\Gamma_{l(r)}$ can only have non-zero elements if both corresponding sites are connected to the left (right) lead. If we index our sites according to their position on the left-right axis (which is sensible considering the geometry of the problem) $\Gamma_{l(r)}$ will have a dense matrix block in the upper left (lower right) corner and be zero everywhere else. Suppose that this matrix block is of dimension $M \times M$ (and that the dimension of the right block is the same). We can then perform the computation in (3.2.2.7) block wise to reduce it to taking a trace of a product of four dense $M \times M$-matrices

$$T_{lr} = \mathrm{Tr}\left[\Gamma_l^{\mathcal{I}\mathcal{I}} G^R_{\mathcal{I}\mathcal{N}} \Gamma_r^{\mathcal{N}\mathcal{N}} G^A_{\mathcal{N}\mathcal{I}}\right], \tag{3.2.2.9}$$

where $\Gamma_l^{\mathcal{I}\mathcal{I}}$ $\left(\Gamma_r^{\mathcal{N}\mathcal{N}}\right)$ denotes the dense $M \times M$ upper left (lower right) corner of $\Gamma_{l(r)}$, $G^R_{\mathcal{I}\mathcal{N}}$ the $M \times M$ upper right corner of G^R and $G^A_{\mathcal{N}\mathcal{I}}$ the $M \times M$ lower left corner of G^A.

We have derived a computationally easy formula (3.2.2.9) for the transmission function from a general formula (3.2.2.7) from the literature [13], [14]. This way of computing the transmission function will give the transmission function between two leads, irrespective of the number of modes within the leads. It only needs a small part of the Green's function to be computed and will also function correctly if the transmission is computed for energies that are not in the transmission band of the lead (this means that the lead is insulating, scattering states do not exist and the imaginary part of the self-energy (and hence the lead velocity operator) is 0), namely that the transmission between two insulating leads as seen from infinity is always

0. When the necessary parts of the Green's function have been derived, (3.2.2.9) will be the most convenient way of computing conductance.

3.3 Example transmission calculations

We now demonstrate how to use the methods of the previous section to compute the transmission function. This will be done using three examples: scattering by a single site with a potential different from the lead chemical potential, the Su-Schrieffer-Heeger model in which the insulator is a one-dimensional wire for which the hopping takes two alternating values and a simple two-dimensional hopping model.

3.3.1 Single site scattering

Our simplest example is that of an insulator that consists of one site and is connected to leads with hopping equal to the hopping within the leads. This model would correspond to an infinite ideal lead with one impurity in the middle. It can be solved analytically and is also treated in [13].

We choose the hopping $t = 1$, the unit of length $a = 1$ and the central site Hamiltonian $H_{1Site} = u + 2$; the band of the leads has total width four and center at 2, so for $u = 0$ we do not have any scattering at all. We restrict our attention to the lead band, since scattering states are only defined if the band admits propagating modes. In this case the S-matrix can be computed analytically, the result is [13]

$$S_{exact} = \begin{pmatrix} \frac{u}{i\hbar v_L - u} & \frac{i\hbar v_L}{i\hbar v_L - u} \\ \frac{i\hbar v_L}{i\hbar v_L - u} & \frac{u}{i\hbar v_L - u} \end{pmatrix}, \tag{3.3.1.1}$$

where v_L is given by (3.1.3.13).

We can also compute the transmission function using methods from the previous section, i.e. by combining (3.1.3.14), the definition of self-energy and (3.2.2.6), after which we either use (3.2.1.2) or (3.2.2.7). It can easily be verified that all three methods give the same result for all u and all energies. It can also be checked that for $u = 0$ the transmission becomes 1 for all energies. In figure 3.2 we show the transmission for $u = 1.5$.

3.3.2 The Su-Schrieffer-Heeger model

The Su-Schrieffer-Heeger model was originally introduced to model properties of polyacetyline [35]. Its Hamiltonian models a one-dimensional chain

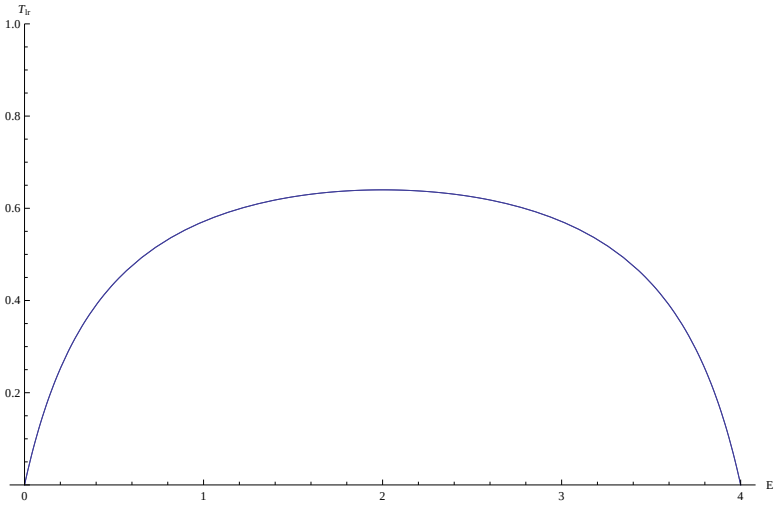

Figure 3.2: The transmission function for two leads separated by a single site scatterer

with alternating nearest-neighbour hoppings t, t' and a chemical potential μ

$$\mathcal{H}_{SSH} = -\left(\left(\sum_{j\in\mathbb{N}_+} \mu c_j^\dagger c_j\right) + \left(\sum_{\substack{j\in\mathbb{N}_+ \\ j \text{ odd}}} t c_j^\dagger c_{j+1} + \text{h.c.}\right) + \left(\sum_{\substack{j\in\mathbb{N}_+ \\ j \text{ even}}} t' c_j^\dagger c_{j+1} + \text{h.c.}\right)\right), \tag{3.3.2.1}$$

where the first site is indexed with 1, the creation and annihilation operators for site j are denoted by $c_j^\dagger, c_j$ and "h.c." means Hermitian conjugate. This Hamiltonian gives rise to a gapped two-band structure (if $t \neq t'$). Furthermore, if the smaller of the two hoppings connects the first to the second site or the final-to-last to the last site, there is a soliton in the band with energy equal to $-\mu$ (i.e. there are 0,1 or 2 of such solitons). These states are however localised and do not contribute to the conductance. The center of the gap is at $-\mu$ and the width of the gap is $2|t - t'|$.

3.3.2.1 Results

We had to scale the hopping within the lead $t_L := -\frac{\hbar^2}{2ma^2}$ and the lead chemical potential such that the system band was contained within the lead band (else the scattering matrix would not be a sensible concept). We connected the leads only to the first and last sites of the insulator with a hopping parameter t_{IL}. We choose $\mu = 0$, $t = 0.8$, $t' = 1.2$, $t_{IL} = 1$, chain length 100 and scaled the leads to have a band from -2.5 to 2.5 (all these numbers are in arbitrary units of energy). The transmission function is shown below in figure 3.3 (the energy E was increased from -2.5 to 2.5 in steps of 0.001 after which the points were joined by straight line segments)

In figure 3.3 we can clearly see a large number of oscillations in the band.

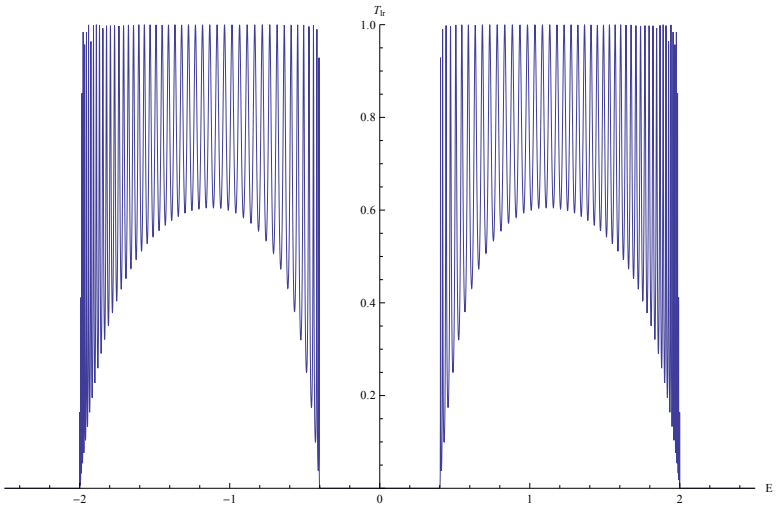

Figure 3.3: The transmission function for the SSH-model. Note that there is indeed a band gap and that the localised states do not contribute to the transmission.

The peaks are at the maximum expected transmission. These oscillations are known as *Fabry-Pérot oscillations* [36] and occur because the bands are not continuous, but consist of discrete points. At the interface between system and lead electrons may reflect, reducing the transmission. There are resonant peaks when a mode in the system is available to a lead electron, at which the transmission is equal to 1. As we will also see in the next example, the precise form of these oscillations depends on to what extend the modes in the lead and in the system are compatible.

3.3.3 Disorder

We can add disorder to the system by adding fluctuations to the chemical potential. Even though this does not seem to make a significant qualitative difference to the band structure, the effects on the transmission function *are* large. This is mainly due to the effects of localisation. We considered again the parameters $\mu = 0$, $t = 0.8$, $t' = 1.2$, $t_{IL} = 1$ and lead band from -2.5 to 2.5. This time however, we added an independent uniformly distributed fluctuation between -0.4 and 0.4 to the chemical potential on each site for 10 chains of length 100 and computed their average transmission function (the energy E was increased from -2.5 to 2.5 in steps of 0.001 after which the points were joined by straight line segments).
We can see in figure 3.4 that the transmission for the disordered system is lower than the transmission of the clean system. If we would have considered

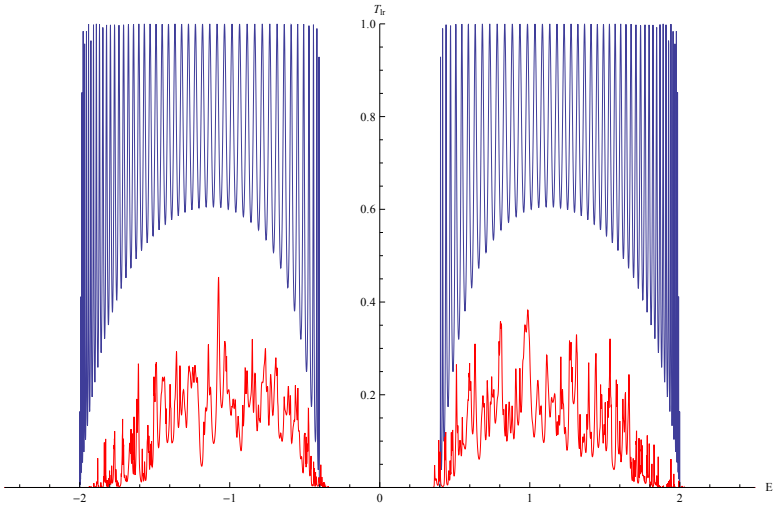

Figure 3.4: The transmission function for the disordered SSH-model (in red). Note that the transmission has become a lot lower than the original transmission (shown in blue). This figure was generated using an ensemble of 10 chains of length 100.

a single chain instead of the ensemble, we would have seen sharper, but higher peaks, but even then only a few peaks would hit 1. The number of peaks would also be smaller than for the clean case. If we look at the spectrum of the chains, it still looks almost the same qualitatively, with two clusters of eigenvalues at the clean-system bands and two states in the gap. This leads to the conclusion that disorder may not be very visible in the spectrum, but that it does hinder transmission significantly through localisation; for more details on this, see appendix B.

3.3.4 Two-dimensional uniform hopping

The next example we look at is two-dimensional: we consider a rectangular lattice of dimension l in the x-direction and dimension w in the y-direction. The lattice spacing is a and we consider a nearest-neighbour hopping model with hopping strength t and chemical potential μ. An example like this is also used in the kwant tutorial, [37], [38]. Because the leads are aligned along the x-direction, we index the sites such that sites in the same column have consecutive indices. The structure of the resulting Hamiltonian is pictured below, with μ in blue and t in red.
Please note that we index the two dimensional lattice as follows: we start by indexing the sites in the first column, starting with the site for which the x- and y-coordinates are the smallest. Then we go to the next smallest y-coordinate etc., until we arrive at the end of the column at site $(1, N_y)$.

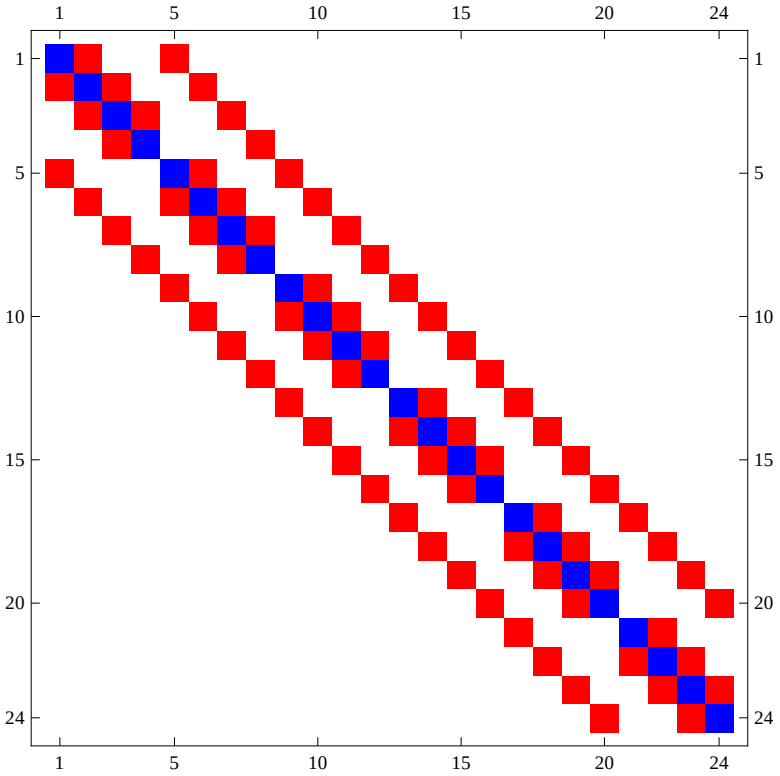

Figure 3.5: A schematic plot of the two-dimensional tight-binding system

We then go to the next column, starting with site $(2, 1)$ and continue. This numbering has the advantage that all sites that are connected to the left lead carry the first N_y indices and all sites connected to the right lead carry the last N_y indices.

3.3.4.1 Two-dimensional ideal leads

In order to consider a two-dimensional system, we need to connect it to two dimensional leads. To do this, we generalize the results from section 3.1.3. We take leads with the same width as the system.

Those ideal leads are described by (2.1.2.3) in two-dimensions, with zero vector potential and constant potential $U(x, y)$, in which we can absorb the band energy. The boundary conditions are Dirichlet boundary conditions on a rectangle $[0, L_x] \times [0, L_y] \subset \mathbb{R}^2$. This problem can be considered to be the superposition of two one-dimensional problems and there is a straight-

forward discretisation [21]

$$-\frac{\hbar^2}{2ma^2}\left(\Psi^k_{j-1}-4\Psi^k_j+\Psi^k_{j+1}+\Psi^{k-1}_j+\Psi^{k+1}_j\right)=(E-U)\Psi^k_j, \quad (3.3.4.1)$$
$$\Psi^k_0=0, \quad \Psi^k_{N_x+1}=0, \quad \Psi^0_j=0, \quad \Psi^{N_y+1}_j=0,$$

where the x-coordinate is indexed by j, the y-coordinate is by k, the lattice spacing is a and we assume $L_x=(N_x+1)a$ and $L_y=(N_y+1)a$.

We introduce $\kappa^x_n=\frac{\pi n}{L_x}$, $\kappa^y_{n'}=\frac{\pi n'}{L_y}$, with $n\in 1,..,N_x$ and $n'\in 1,..,N_y$. The eigenfunctions solving (3.3.4.1) are given by

$$\Psi^k_j(\kappa^x_n,\kappa^y_{n'})=\sqrt{\frac{4}{L_xL_y}}\sin(\kappa^x_n ja)\sin(\kappa^y_{n'}ka). \quad (3.3.4.2)$$

The dispersion relation is

$$E(\kappa^x_n,\kappa^y_{n'})=U-\frac{\hbar^2}{ma^2}\left(\cos(\kappa^x_n a)+\cos(\kappa^y_{n'}a)-2\right) \quad (3.3.4.3)$$

and the retarded Green's function is given by

$$G(\epsilon,j,k;j',k')=\sum_{n'=1}^{N_y}\sum_{n=1}^{N_x}\frac{\Psi^k_j(\kappa^x_n,\kappa^y_{n'})\left(\Psi^{k'}_{j'}(\kappa^x_n,\kappa^y_{n'})\right)^*}{\epsilon-E(\kappa^x_n,\kappa^y_{n'})+i\eta} \quad (3.3.4.4)$$

$$=\frac{4}{L_xL_y}\sum_{n'=1}^{N_y}\sum_{n=1}^{N_x}\frac{\sin(\kappa^x_n ja)\sin(\kappa^y_{n'}ka)\sin(\kappa^x_n j'a)\sin(\kappa^y_{n'}k'a)}{\epsilon-U+\frac{\hbar^2}{ma^2}\left(\cos(\kappa^x_n a)+\cos(\kappa^y_{n'}a)-2\right)+i\eta}. \quad (3.3.4.5)$$

We can compute the Green's function using the results from section 3.1.3: we keep L_y fixed and take L_x to infinity. We introduce $t=\frac{-\hbar^2}{2ma^2}$ and $\bar{\epsilon}(n'):=2\left(\frac{ma^2(\epsilon-U)}{\hbar^2}-1+\cos(\kappa^y_{n'}a)\right)$ such that all the results carry over from section 3.1.3 if we take $\bar{\epsilon}\to\bar{\epsilon}(n')$. We can express (3.3.4.5) in terms of (3.1.3.10)

$$G^{2D}(\epsilon,j,k;j',k')=\frac{2}{L_y}\sum_{n'=1}^{N_y}\sin(\kappa^y_{n'}ka)\sin(\kappa^y_{n'}k'a)G^{1D}(\epsilon(n'),j,j'). \quad (3.3.4.6)$$

We will be mainly interested in the case where $j=j'=1$, so we get

$$G^{2D}(\epsilon,1,k;1,k')=\frac{2}{L_y}\sum_{n'=1}^{N_y}\sin(\kappa^y_{n'}ka)\sin(\kappa^y_{n'}k'a)\frac{-1}{ta}e^{\pm i\kappa_L(n')a}, \quad (3.3.4.7)$$

where $\kappa_L(n')=\frac{-i}{a}\log(z_+(\bar{\epsilon}(n'))$. This means that we can compute the exact two-dimensional lead Green's function for systems that are not too wide, after which the methods of section 3.2.2 can be applied.

3.3.4.2 Results

The two-dimensional systems under consideration have as many bands as their width. These bands all cover a range of energies. We can estimate the transmission at a given energy by drawing a line of constant energy through the band structure and counting how many bands are "hit" by this line of constant energy. This is then the expected transmission.

Suppose we choose an insulator of width 4, with $\mu = 0$ and $t = 1$. We first derive its (bulk) band structure, which is done by applying periodic boundary conditions in the x-direction and then taking the limit where the period length goes to infinity. We can then perform a Fourier transform on the x-coordinate and obtain 4 continuous bands shown in figure 3.6

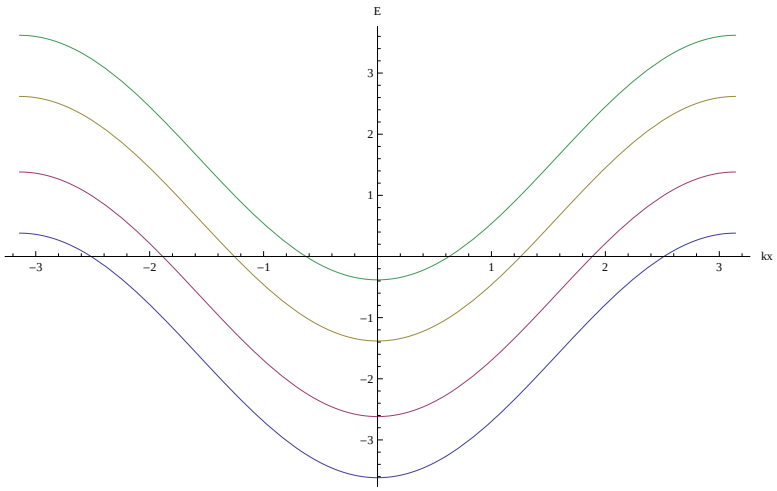

Figure 3.6: The bulk band structure of a 4 mode 2D wire

Looking at figure 3.6, we expect a transmission function consisting of plateaus at 1, 2, 3 and 4, symmetric around 0. No Fabry-Pérot oscillations are expected, since this is a rather fine-tuned system in which the leads and centre are all exactly the same, which means our system is in fact a periodic infinite lead of width 4. As shown in figure 3.7, our expectations agree with the result[7].

3.3.4.3 Disorder

By adding some random fluctuations to the local chemical potential we can add disorder to the system. It is expected that the disorder will not change

[7]Which is also consistent with the results in [37], [38]

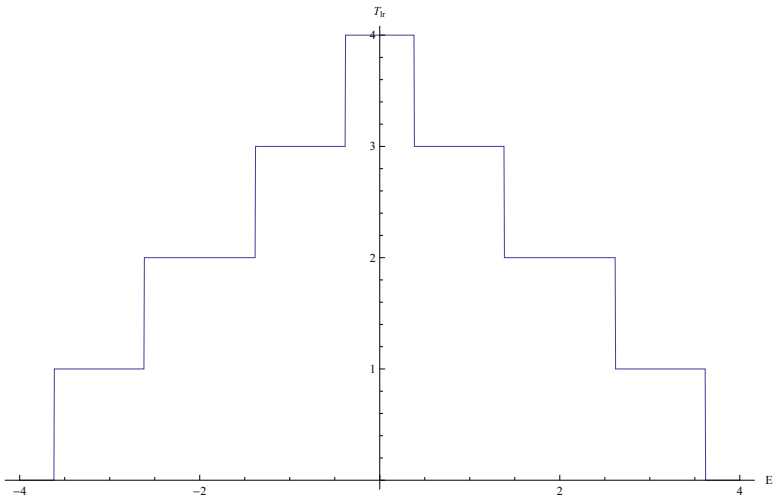

Figure 3.7: The transmission of a 4 mode 2D wire as a function of energy

the spectrum significantly (though it can become less regular), but that localisation effects will strongly reduce the values of the transmission function.

For our test of the effects of disorder we used the same system as the system used in figure 3.7. The disorder was modelled by adding to the local chemical potential fluctuations which were independent and uniformly distributed between -0.5 and 0.5. The ensemble average of 10 independent systems was taken. In order to limit the computational time, the step size in the energy was 10 times as large as the step size used in figure 3.7. The result is shown in figure 3.8.
In figure 3.8 we can see that the transmission function is on average much smaller, especially in the energy range for which the clean system has the highest transmission rates. This reduced transmission rate is due to localisation effects. If we were to increase the disorder, this would lead to even smaller transmissions; if we for example take fluctuations that are uniformly distributed between -2 and 2, the transmission becomes negligible, which is consistent with appendix B.

3.4 The lead Green's function for arbitrary tight-binding models

It turns out that in order to study the transmission of the topological states of matter discussed in section 5.1, we have to further generalize our ideal lead computations. Reasons for the need of this generalization will be given in the next subsection. The rest of this section contains a derivation of a

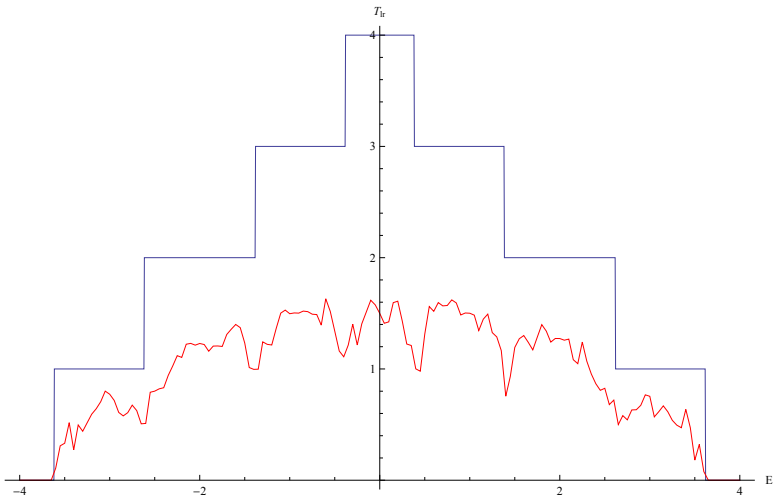

Figure 3.8: The transmission of a 4 mode 2D disordered wire as a function of energy

(mostly) stable way of computing these more general lead Green's functions numerically, following [14].

3.4.1 Why ideal metals can not always be used as leads

The tight-binding Hamiltonians that realize topological insulators will in general not be similar to the systems considered in section 3.3. For instance, one of the features of the Hamiltonian in section 5.2 is that it contains a next-nearest neighbour hopping term.

When we want to couple our leads to the system, we have to explicitly specify the way in which sites of the lead are coupled to sites of the insulator. If the leads and the insulator have the same global structure, as in section 3.3, it is clear which lead-site should be connected to which insulator site and the coupling matrices can be defined easily.

However, if the structure of the couplings is different between insulator and leads, there is some ambiguity in how to couple the different parts of the system. We could of course resolve that ambiguity "by hand", that is, we could just pick some convenient coupling matrix and proceed as usual. However, the choice we make does not have to (and likely will not) correspond to the effective coupling that would be realized in an experimental situation.

Since it would be very non-trivial to derive what the "physical" coupling for a general system would be, we strive to obtain a situation comparable

to that in section 3.3 instead: we would like to choose the lead to have the same structure as the insulator. We do this by using the same Hamiltonian for lead and insulator, but to guarantee the lead to be transmitting, we will shift the chemical potential of the lead such that for the energy range of interest we are in the lead band, which is conducting.

3.4.2 Derivation of the lead Green's function

We are now faced with the task of computing the semi-infinite lead Green's function for arbitrary tight-binding models. This problem has been solved many times in literature (see for example [39, 40, 41, 42, 14] and references therein). We will follow the approach used in [14], to which we refer the reader for additional details.

Suppose we have an *infinite* lead, consisting of a periodic repetition of unit cells. It's Hamiltonian is block three-diagonal: on the block diagonal there is an infinite repetition of the unit cell Hamiltonian H_0, the upper block diagonal consists of an infinite repetition of the hopping Hamiltonian H_1, enabling electrons to hop from one unit cell to its left neighbour and the hopping to the right is governed by an infinite repetition of $H_{-1} = H_1^\dagger$ on the lower block diagonal.

Because the infinite lead is periodic, we can use Bloch's theorem (see section 2.1.2) to write the eigenfunctions as

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

where j is a unit cell index, $k \in [-\pi, \pi]$ and $\phi_{n,k}$ and eigenvector of $H_0 + H_1 e^{ik} + H_{-1} e^{-ik}$ (with eigenvalue $E_{n,k}$). For fixed k, the eigenvectors $\phi_{n,k}$ are a complete orthonormal system.

We remark that we can also interpret the infinite lead as two semi-infinite leads that are connected with a hopping matrix: it then follows from Dyson's equation (4.1.1.3) that [14]

$$G(j, j+1) = g(j+1, j+1) H_{-1} G(j, j), \tag{3.4.2.1}$$

where $g(j,k)$ is the (retarded) semi-infinite (right-)lead Green's function and $G(j,k)$ the (retarded) infinite lead Green's function between sites j and k.

As in section 3.1.3, we write the Green's function as a contour integral which we then attempt to solve. Proceeding analogously to (3.1.3.4)-(3.1.3.6) and scaling the length such that the unit cell spacing $a = 1$, we obtain [14]

$$G(\epsilon, j, j') = \frac{-i}{2\pi} \oint_{|z|=1} z^{j-j'} \left(\left((\epsilon + i\eta)\mathbf{1} - H_0 \right) z - H_1 z^2 - H_{-1} \right)^{-1} \mathrm{d}z, \tag{3.4.2.2}$$

where ϵ and η are defined as in section 3.1.3 and $\hat{\mathbf{1}}$ is an appropriate identity matrix.

In order to invoke Cauchy's theorem for the contour integral (3.4.2.2) we should identify which quadratic eigenvalues λ_n and -vectors $\mathbf{u}_n$ solve the equation [14]

$$\left(\left((\epsilon + i\eta)\hat{\mathbf{1}} - H_0\right)\lambda_n - H_1\lambda_n^2 - H_{-1}\right)\mathbf{u}_n = 0. \tag{3.4.2.3}$$

As in section 3.1.3, we intend to solve 3.4.2.3 without giving η a non-zero numerical value: instead, we can analyse the effect of an infinitesimal η by looking at the eigenvalues and vectors for the system with $\eta = 0$ [14].

In practice, we solve (3.4.2.3) by writing it as a linear generalized eigenproblem of twice the size, that is, we introduce $\mathbf{v_n} := \lambda_n \mathbf{u}_n$ and solve the system[8]

$$\begin{pmatrix} \hat{\mathbf{0}} & \hat{\mathbf{1}} \\ -H_{-1} & \epsilon\hat{\mathbf{1}} - H_0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_n \\ \mathbf{v}_n \end{pmatrix} = \lambda_n \begin{pmatrix} \hat{\mathbf{1}} & \hat{\mathbf{0}} \\ \hat{\mathbf{0}} & H_1 \end{pmatrix} \begin{pmatrix} \mathbf{u}_n \\ \mathbf{v}_n \end{pmatrix}. \tag{3.4.2.4}$$

Having solved this eigenproblem,[9] we compute the velocities (see (3.1.3.13)) for eigenvectors with absolute eigenvalue 1 and classify the eigenvalues and -vectors into two equally sized groups [14]:

- Left-goers, with $|\lambda_n| > 1$ (left-decayers) or $|\lambda_n| = 1$ and negative velocity (left-movers).
- Right-goers, with $|\lambda_n| < 1$ (right-decayers) or $|\lambda_n| = 1$ and positive velocity (right-movers).

It can be shown [14] that eigenvalues corresponding to right-goers are inside the unit circle for infinitesimal η and left-goers are outside. We denote the matrix with columns consisting of right-going eigenvectors by $U_<$, the corresponding diagonal matrix of eigenvalues by $\Lambda_<$ and $V_< := U_<\Lambda_<$. For the right-goers these matrices are denoted by $U_>$, $\Lambda_>$ and $V_>$ respectively. These matrices will be square and their size is denoted by N. We also define

$$U_1 := \begin{pmatrix} U_< \\ V_< \end{pmatrix},$$
$$U := \begin{pmatrix} U_< & U_> \\ V_< & V_> \end{pmatrix}.$$

[8] The substitution $\mathbf{v_n} := \lambda_n \mathbf{u}_n$ is similar to a substitution often used to turn a second order differential equation into a system of first order differential equations, see f.e. [43] and [21].

[9] Algorithms to solve this generalized eigenproblem are given in [44].

Assuming without proof that $U_<$ is invertible,[10] the contour integral (3.1.3.6) can be computed [14]

$$G(\epsilon, j+1, j) = U_< \Lambda_< U_<^{-1} G(\epsilon, j, j). \tag{3.4.2.5}$$

Combining (3.4.2.5) and (3.4.2.1) yields the lead Green's function $g(\epsilon)$ and self-energy $\Sigma(\epsilon)$

$$g(\epsilon) H_{-1} = U_< \Lambda_< U_<^{-1}, \tag{3.4.2.6}$$

$$\Sigma(\epsilon) = H_1 U_< \Lambda_< U_<^{-1}. \tag{3.4.2.7}$$

3.4.3 Towards a stable computation of the lead Green's function

The algorithm sketched above does yield the lead Green's function for very general choices of Hamiltonians, but there are still some situations in which is suffers from stability problems. We now identify the causes of these problems and attempt to find solutions to them.

There are three kind of problems that can destabilize the algorithm:

1. The generalized eigenvalues λ_n solving (3.4.2.4) may be degenerate.
2. The velocities of the non-decaying eigenvectors may be very close to 0.
3. Even if $U_<$ is invertible in exact arithmetic[11] its condition number (see appendix C) may become very high (see [14] for an example).

The degeneracy of eigenvalues can be problematic, since then the eigenvectors will not be uniquely defined any more. For decaying eigenvalues, this is no problem, since the eigenvectors will still be classified correctly into right- or left-goers and the final result (3.4.2.6) will not change. If the degenerate eigenvalues do have absolute value 1, the computation of the velocities becomes ill-defined. However, this is a rare situation and if the degenerate eigenvalues are both left-going or both right-going, this problem can be handled [14].

Velocities that are close to zero are an issue however: because there will

[10] This can not be proven in general; if $U_<$ happens to be singular we are out of luck, but fortunately this does not usually happen in practice.

[11] If it is not invertible, then the derivation of 3.4.2.5 breaks down as well as subsequent steps; the solutions discussed later in this section will not help and become ill-defined themselves.

always be numerical errors in the computed velocities, this leads to misidentification of left- and right-going modes. In this way $U_<$ and $\Lambda_<$ will consist of the wrong eigenvalues and -vectors and the final answer (3.4.2.6) becomes plain incorrect.

To prevent this from occurring, we should make sure that either these very small velocities are computed to high enough accuracy or we should pick our energy such that no velocity is too small. These two approaches are of course two sides of the same medal, since higher accuracy will make larger energy ranges permissible. We also remark that if this occurs small shifts in energy will not make a difference, since the velocities will depend on the energy smoothly.

If $U_<$ is invertible but ill-conditioned, (3.4.2.5) still holds, but it will not make sense any more to compute the lead Green's function, because of numerical inaccuracies. $U_<$ can either be ill-conditioned because U itself is ill-conditioned or because we take a submatrix. Following [14], we rewrite (3.4.2.6) to depend on submatrices (and their inverses) of orthogonal matrices. In principle the orthogonality of a matrix does not prove anything about the condition numbers of their submatrices, but in practice we do expect the conditioning of the submatrices to improve.

Let B be any invertible $N \times N$ matrix and define $U_1' := U_1 B$, $U_<' := U_< B$ and $V_<' := V_< B$. Then

$$\begin{aligned} g(\epsilon) H_{-1} &= V_< U_<^{-1} \\ &= V_< B B^{-1} U_<^{-1} \\ &= V_<' U_<'^{-1}. \end{aligned}$$

Our goal is now to construct an invertible matrix B, such that the condition number of $U_<'$ is lower. The idea is to focus on bringing the condition number of U to 1 and then use the transformation that does this as our matrix B.

We assume H_1 to be invertible[12]. Define the matrix A corresponding to (3.4.2.4)

$$A := \begin{pmatrix} \hat{\mathbf{0}} & \hat{\mathbf{1}} \\ -H_1^{-1} H_{-1} & H_1^{-1} \left(\epsilon \hat{\mathbf{1}} - H_0 \right) \end{pmatrix}.$$

Of course, the columns of U are the eigenvectors of A. Suppose we only wanted to find the eigenvalues of A. A possible approach would then be to find the Schur decomposition[13] of A. The Schur decomposition of A consists of a unitary matrix Q and an upper triangular matrix T such that

[12] For singular H_1, see [14]

[13] See for example [44]

$A = QTQ^\dagger$.

Since the spectra of A and T are the same, all eigenvalues of A appear on the diagonal. Because Q is unitary, its condition number subordinate to the 2-norm,[14] $\kappa_2(Q) = 1$ [44], [45].

We furthermore note that we can adjust T and Q such that the eigenvalues of A appear in any desired order on the diagonal of T and that this reordering is computationally cheap compared to computing the original Schur decomposition[15] [44].

Now assume the first half of the eigenvalues on the diagonal of T correspond to right-going modes and that there are no eigenvalues that are degenerate and both right- and left-going[16], then there exist block matrices

$$X = \begin{pmatrix} X_1 & \hat{\mathbf{0}} \\ \hat{\mathbf{0}} & X_2 \end{pmatrix}, \; Y = \begin{pmatrix} \hat{\mathbf{1}} & Y_{12} \\ \hat{\mathbf{0}} & \hat{\mathbf{1}} \end{pmatrix},$$

such that [44], [14]

$$X^{-1}Y^{-1}Q^\dagger AQYX = \begin{pmatrix} \Lambda_< & \hat{\mathbf{0}} \\ \hat{\mathbf{0}} & \Lambda_> \end{pmatrix}. \tag{3.4.3.1}$$

Since (3.4.3.1) is nothing else but the solution of (3.4.2.4), we can identify the transformation $Q = Y^{-1}X^{-1}U$ that transforms the basis given by the columns of U into an orthonormal one. It now follows that

$$g(\epsilon)H_{-1} = Q_{21}Q_{11}^{-1}, \tag{3.4.3.2}$$

where the indices denote matrix blocks of size $N \times N$.

We remark again that the use of the Schur decomposition does not guarantee that the condition number of the block Q_{11} is close to 1 and examples can easily be given in which a subblock of a unitary matrix is singular. It does however seem reasonable to assume that the conditioning of the blocks of Q is better than the conditioning of the blocks of U.

In summary, we have identified some causes for instability of the numerical computation of the Green's function for an arbitrary lead. The solutions proposed in this section may help stabilize the algorithm in some cases, but there remain situations for which the lead Green's can not be computed (reliably).

[14] See appendix C for definitions.
[15] In Matlab this can be done using the ordschur() routine [46].
[16] so $\mathrm{diag}(\Lambda_<) \cap \mathrm{diag}(\Lambda_>) = \emptyset$

3.4.4 A pragmatic approach to lead Green's function computational stability

The numerical computation of the lead Green's function as described above is not unconditionally stable and it is not obvious if and how this could be fixed. In many cases we can however compute it without trouble and we can use physical arguments to justify restricting our attention to the cases in which the lead Green's function is computable with our algorithm.

The motivation for computing the lead Green's function for a general Hamiltonian was to couple a system to a lead that only differed in the value of the chemical potential. This is both more physical and less ad-hoc then specifying some coupling to an ideal lead ourselves. A shift of the chemical potential is still not entirely physical, but does not invalidate the computations, because the leads function only as source and sink of electrons.

A way to speed-up and stabilize computations consists of removing the energy dependence of the leads. This means we really fix the sum of the chemical potential and the energy within the leads. Even though it could be argued that the computational system now bears less resemblance to an experimental situation, it is important to realize that our goal is not to resemble the experimental situation as closely as possible, but only as closely as needed to obtain comparable results.

The step we take here can either be interpreted as "attaching a different material as lead at each different energy" or as "keeping the density of states of our sources and sinks constant at all energies". Regardless of the interpretation, we make the leads more abstract here, giving computational advantages while not giving physical disadvantages.

Chapter 4

Computational methods for the transmission function

As can be seen from the Fisher-Lee relation (3.2.1.3) and the compact formula for the transmission function (3.2.2.9), we do not need to know the full Green's function in order to compute the conductance. Furthermore, the tight-binding Hamiltonian is a sparse matrix. These properties can be exploited to compute the transmission function efficiently.

In this chapter we compare three of such methods: the recursive Green's function method [47], [48], [49], (a standard way of doing this in physics), a sparse direct method (an optimised variant of LU-decomposition (see appendix C) as provided by MATLAB [46]) and an iterative method, which uses the generalised minimal residuals algorithm (GMRES) with a preconditioner (see appendix C and [50]).

4.1 The recursive Green's function algorithm

The recursive Green's function algorithm is a popular way to compute transmission functions. It is based on physical considerations and most suitable for the computation of transmission through systems that are long in the direction of the leads, but not very wide.

4.1.1 Derivation of the algorithm

The basic idea of the recursive Green's function algorithm is to start with a part of the system for which we know the Green's function as long as this part is not connected to the rest of the system (i.e. one of the leads). We then consider the part of the system that is connected to the known part (i.e. a slice of the insulator), compute the Green's function for this slice if it were disconnected and then connect the two disconnected parts for which we

know the Green's function and coupling, treating the coupling as a perturbation using the Dyson equation derived below. We do however remark that the Dyson equation is exact and that we will not use any approximations in solving it, thus incorporating the coupling in an exact manner.

This will yield the relevant parts of the Green's function for the connected system, after which we add another slice. Our derivation in section 4.1.1.2 follows [14], though the algorithm there starts at the right lead, while we will start at the left lead. In order to give a more precise description of this algorithm, we will first derive the central equation of the recursive Green's function method, the Dyson equation.

4.1.1.1 The Dyson equation

The Dyson equation relates Green's functions of non-interacting systems in quantum field theory to Green's functions for systems which do interact, see for example [28]. We will specifically derive it for matrices here, since that is the case for which we need it.

Let $\hat{H}_0$ be a square matrix (i.e. an insulator Hamiltonian into which the lead self-energy is already incorporated). We denote the matrix $E\hat{\mathbf{1}} - \hat{H}_0$[1] by G_0^{inv}, assume it is invertible and denote its inverse by G_0. Now define other matrices V and $\hat{H} = \hat{H}_0 + V$ and G^{inv}, G analogously. We then have

$$\begin{aligned} G_0^{inv} &= E\hat{\mathbf{1}} - \hat{H}_0 \\ &= E\hat{\mathbf{1}} - \hat{H} + V \\ &= G^{inv} + V. \end{aligned} \tag{4.1.1.1}$$

Furthermore, we remark that we are dealing with matrices here, so all inverses are both left- and right inverses. We get

$$\begin{aligned} GG^{inv} &= G_0 G_0^{inv} \\ &= G_0 \left(G^{inv} + V\right) \\ &= G_0 \left(G^{inv} + VGG^{inv}\right) \\ &= \left(G_0 + G_0 VG\right) G^{inv}, \end{aligned} \tag{4.1.1.2}$$

from which the Dyson equation

$$G = G_0 + G_0 VG \tag{4.1.1.3}$$

follows immediately. Another form of the Dyson equation, which can be computed analogously, is given by

$$G = G_0 + GVG_0 \tag{4.1.1.4}$$

[1] where $E \in \mathbb{R}$ and $\hat{\mathbf{1}}$ an appropriate identity matrix

4.1.1.2 The block-wise computation of the Green's function

We index our system such that we can write our Hamiltonian as a block tridiagonal matrix, with the upper left block corresponding to the first slice of our insulator, connected to the left lead and the lower right block corresponding to the last slice, connected to the right lead (here a "slice" is a subsystem with dimension one lower than the studied system).

Suppose that there are a total of $N+1$ slices. We can then define the following sequence of subsystems: subsystem $L(0)$ consists only of the left lead. Subsystem $L(1)$ consists of the left lead and the first slice of the insulator. Subsystem $L(i)$, with $1 \leq i \leq N$ consists of the left lead and the first i slices of the insulator and subsystem $L(N+1)$ consists of the full system, including the right lead.

We assume that we know the Green's function for $L(0)$ (see section 3.1.3). We furthermore assume that we can invert square matrices with the same dimensions as the number of sites per slice easily. We define $\hat{H}_{L(i),0}$ to be the Hamiltonian of a system consisting of $L(i-1)$ and the *disconnected* slice i (note that this is *not* a subsystem of our system under consideration, but it is the sum of two subsystems). We let $\hat{H}_{L(i)}$ be the Hamiltonian of $L(i)$ and $V := \hat{H}_{L(i)} - \hat{H}_{L(i),0}$. The corresponding Green's functions will be referred to as $G_{L(i),0}, G_{L(i)}$. We remark that the ith row and column of $G_{L(i),0}$ vanishes for all off-diagonal blocks and that the only non-zero blocks of V are $V^{i-1,i}, V^{i,i-1}$.

We want to use the Dyson equation to compute blocks situated at the top row. In order to do this we need to apply the equation for the following blocks

$$G^{i,i}_{L(i)} = G^{i,i}_{L(i),0} + G^{i,i}_{L(i),0} V^{i,i-1} G^{i-1,i}_{L(i)}, \tag{4.1.1.5}$$

$$G^{i-1,i}_{L(i)} = G^{i-1,i}_{L(i),0} + G^{i-1,i-1}_{L(i),0} V^{i-1,i} G^{i,i}_{L(i)}, \tag{4.1.1.6}$$

$$G^{1,i}_{L(i)} = G^{1,i}_{L(i),0} + G^{1,i-1}_{L(i),0} V^{i-1,i} G^{i,i}_{L(i)}. \tag{4.1.1.7}$$

We use the fact that $G^{i-1,i}_{L(i),0}$ vanishes and substitute (4.1.1.6) into (4.1.1.5) to obtain

$$G^{i,i}_{L(i)} = G^{i,i}_{L(i),0} + G^{i,i}_{L(i),0} V^{i,i-1} G^{i-1,i-1}_{L(i),0} V^{i-1,i} G^{i,i}_{L(i)}. \tag{4.1.1.8}$$

We now note that (4.1.1.8) is solved by

$$G^{i,i}_{L(i)} = \left(\left(G^{i,i}_{L(i),0} \right)^{-1} - V^{i,i-1} G^{i-1,i-1}_{L(i),0} V^{i-1,i} \right)^{-1} \tag{4.1.1.9}$$

and for $i > 1$ (4.1.1.7) simplifies to [2]

$$G^{1,i}_{L(i)} = G^{1,i-1}_{L(i),0} V^{i-1,i} G^{i,i}_{L(i)}. \tag{4.1.1.10}$$

The equations (4.1.1.9) and (4.1.1.10) enable us to compute the relevant blocks of the Green's function, provided that we can compute some quantities for the disconnected system. It is easy to see that they are given by

$$G^{j,k}_{L(i),0} = \begin{cases} G^{j,k}_{L(i-1)}, & \text{if} \quad j < i \text{ and } k < i, \\ \left(E - \hat{H}_i + \sqrt{-1}\eta\right)^{-1}, & \text{if} \quad j = k = i, \\ 0, & \text{else}, \end{cases} \tag{4.1.1.11}$$

where the $\hat{H}_i$ is the Hamiltonian of the isolated slice i.[3] This gives us the recursion

$$G^{i,i}_{L(i)} = \left(E - \hat{H}_i - V^{i,i-1} G^{i-1,i-1}_{L(i-1)} V^{i-1,i}\right)^{-1} \tag{4.1.1.12}$$

$$G^{1,i}_{L(i)} = G^{1,i-1}_{L(i-1)} V^{i-1,i} G^{i,i}_{L(i)}. \tag{4.1.1.13}$$

We remark that we can leave the infinitesimal imaginary part out of (4.1.1.12) (see section 3.2.2). The recursion (4.1.1.12), (4.1.1.13) is started by setting $G^{0,0}_{L(0)} = \mathcal{G}^{0,0}_L$, where $\mathcal{G}_L$ is the Green's function for the isolated left lead and $V^{0,1} = \tau_l(0,1), V^{1,0} = \tau_l^\dagger(1,0)$. In the first step, where $i = 1$, we do not use (4.1.1.13).

From $i = 2$ until $i = N$ the recursion is straightforward. We note that $G^{1,N}_{L(N)}$ is obtained this way, but that we still need to connect the system to the right lead, which means that we should in fact compute $G^{1,N}_{L(N+1)}$! We can compute this using (4.1.1.9), (4.1.1.13), (4.1.1.4) and known properties of right lead and coupling, resulting in the following final step

$$\begin{aligned} G^{N+1,N+1}_{L(N+1)} &= \left(\left(\mathcal{G}^{N+1,N+1}_R\right)^{-1} - \tau_r^\dagger(N+1,N) G^{N,N}_{L(N)} \tau_r(N,N+1)\right)^{-1}, \\ G^{1,N+1}_{L(N+1)} &= G^{1,N}_{L(N)} \tau_r(N,N+1) G^{N+1,N+1}_{L(N+1)}, \\ G^{1,N}_{L(N+1)} &= G^{1,N}_L(N) + G^{1,N+1}_{L(N+1)} \tau_r^\dagger(N+1,N) G^{N,N}_{L(N)}, \end{aligned} \tag{4.1.1.14}$$

where $\mathcal{G}_R$ is the Green's function for the isolated right lead. Having computed $G^{1,N}_{L(N+1)}$ with (4.1.1.12), (4.1.1.13) and (4.1.1.14), we can apply (3.2.2.9) to obtain the transmission function without having to compute the inverse of the full matrix.

[2] For i=1 we should compute this block using (4.1.1.9)

[3] In 4.1.1.11 all the i's are indices.

4.1.2 Computational cost of the recursive Green's function algorithm

Consider a d-dimensional insulator ($d \in \{1, 2, 3\}$). We assume it can be divided in N square slices with side M (i.e. the slices have M^{d-1} sites). We estimate the computational cost in terms of the asymptotic behaviour of the number of basic computational operations as $M, N \to \infty$.

Even though the results obtained this way will give a far from complete picture, it is still useful to give an idea of how the program will scale to large insulators, the analysis is easy and it is a very common way to estimate the performance of algorithms.

We want to compare the recursive algorithm with the cost of just applying (3.2.2.6). To apply (3.2.2.6) directly, we should invert a matrix with size $\left(NM^{d-1}\right)^2$. There are multiple ways (such as LU-decomposition or Gauss-Jordan elimination) in which the full inverse can be computed, with computational cost of $\mathcal{O}\left(\left(NM^{d-1}\right)^3\right)$.

The recursive Green's function algorithm consists of an initialisation step, the actual recursion and a finalisation step. Only the actual recursion (4.1.1.12),(4.1.1.13) depends on the system length.

We first analyse the cost of one recursion step. This means we should solve (4.1.1.12),(4.1.1.13) once. For (4.1.1.12) we have to do two matrix multiplications, two matrix additions and one inversion; for (4.1.1.13) we should perform two matrix multiplications. All matrices are dense matrices with dimensions $\left(M^{d-1}\right) \times \left(M^{d-1}\right)$ (this is due to the fact that all slices are strongly coupled to the rest of the system). We assume multiplications and inversion to be $\mathcal{O}\left(\left(M^{d-1}\right)^3\right)$ operations (see appendix C); the additions scale as $\mathcal{O}\left(\left(M^{d-1}\right)^2\right)$.

This means that the total cost per recursion step is $\mathcal{O}\left(\left(M^{d-1}\right)^3\right)$. Since there are N steps, the recursion scales as $\mathcal{O}\left(NM^{3(d-1)}\right)$. Initialisation and finalisation do not depend on the system length. For this the lead Green's function should be computed; this has to be done for any method and will generally cost $\mathcal{O}\left(M^{3(d-1)}\right)$ (see sections 3.3.4.1 and 3.4). Having computed this lead Green's function, a few multiplications and inversions should still be done, also costing $\mathcal{O}\left(M^{3(d-1)}\right)$.

The total cost will hence be given by $\mathcal{O}\left(N\left(M^{d-1}\right)^3\right)$. The cost dependence on slice size has not improved, but there is a significant improvement

on the length dependence: this used to depend on the length cubed and has become linear.

In principle it is possible to parallelize the algorithm, though it should of course be reformulated quite a bit first [51]. In our implementation we will however not pursue this route, since there are usually many different transmissions to be computed in one run and trivial parallelisation (assigning different transmission to different computational units) will do.

4.1.3 Stability of the recursive Green's function algorithm

A look at the main recursion and appendix C make it seem that the recursive Green's function algorithm is unlikely to be unconditionally stable. The large amounts of multiplications and inversions will introduce errors and all intermediate results effect later (intermediate and final) results. It is furthermore not straightforward to apply perturbation theory to fully analyse the algorithm.

Another property of the recursive Green's function algorithm is that only the block we are actually interested in is computed, whereas in the other methods discussed in this chapter a number of columns of the full Green's function is computed. The advantage of this is that we do not compute quantities we do not need for the final answer, but it also has the disadvantage that we can not just compute the residual to obtain an error estimate, as opposed to the more standard methods where we can easily check if applying the inverse Green's function to the computed Green's function columns yield a result that is somehow close to columns of the identity matrix or not.

We do not have to dismiss the recursive Green's function algorithm because of this, however. If problems are to occur in the execution of the algorithm, it is likely that some matrix elements become very big. When this happens, the "inv()" routine in Matlab (used in our implementation) should complain about condition numbers that are estimated to be very high. This particular warning however is extremely rare, which can give us hope that the results are reliable (but it does not guarantee it).

A better way to ascertain stability is to compare the results to those obtained using a different algorithm, preferably one for which we know it is stable. Due to the computational cost, this should not be done in all cases, but a few tests can also reinforce our confidence in the recursive Green's function algorithm. A good candidate for a "comparison algorithm" is the LU decomposition discussed below: the stability properties can be found in the literature and are straightforward to evaluate for a given problem. We did this stability test for some systems discussed throughout the thesis and

the differences in the computed transmission functions were always negligibly small.

Another reason to assume the algorithm to be fairly stable is the fact that it is often used in physics and the results seem to be sensible [14]. In our computations with this algorithm we also found the results to be physically justifiable.

To come up with a more quantitative check of the RGF algorithm, we reconsider the system in section 3.3.4. We pick a width of 10, $t = 1$, $\mu = 0$, energy $E = 0$ and leads identical to the system. For this system we know the conductance to be exactly 10 for any system length.[4] We compute the self-energy using the methods from section 3.4 and show the absolute error as a function of system length in table 4.1.

Length	Absolute error
1	1.78×10^{-15}
10	7.11×10^{-15}
100	4.44×10^{-14}
1000	3.41×10^{-13}
10000	1.74×10^{-12}
100000	4.20×10^{-12}

Table 4.1: Absolute error as a function of length for the computed transmission of 10 connected infinite leads, using a Matlab implementation of the RGF algorithm, combined with the self-energy computation described in section 3.4.

We can see in table 4.1 that the error does depend on the system length, but even for very large systems, with up to a million sites, the error remains very small (compared to the transmission). We furthermore remark that the error for a system of length one is in fact equal to the smallest representable difference between the double 10 and the next representable double.

All in all, we can say that the recursive Green's function formalism is usable, but we should of course check if our results make any sense, in which case we can pragmatically assume that the computation was apparently stable. If we are in doubt, we should re-do the particular computation with an alternative method.

[4] In this case the distinction between "leads" and "central region" is arbitrary and we are in fact computing the conductance of (a part of) an infinite ideal lead of width 10, at an energy and chemical potential such that all modes contribute to the conductance.

4.2 Direct column-wise solution

We need only a few elements in the full Green's function and the recursive Green's function is a way of computing these elements and too many unnecessary elements in the full Green's function. An alternative way of computing only a small part of the matrix is by only considering the relevant columns in (3.2.2.6). We then solve the equations for the columns of the Green's function using either a direct or an iterative method.

As in section 4.1.2, we consider a system with length N and width M in dimension $d \in \{1, 2, 3\}$. In order to use (3.2.2.9), we have to solve M^{d-1} equations of the form

$$A\mathbf{x} = \mathbf{b}, \tag{4.2.0.1}$$

where $A = \left((E + i\eta)\hat{\mathbf{1}} - \hat{H}_i - \Sigma\right)$ is a $NM^{d-1} \times NM^{d-1}$ matrix, $\mathbf{x}$ is an unknown vector and $\mathbf{b}$ is a given (standard basis) vector. As stated in appendix C, the general problem $\hat{A}\mathbf{x} = \mathbf{b}$ is at most of order $\mathcal{O}\left(\left(NM^{d-1}\right)^2\right)$ and at least of order $\mathcal{O}\left(\left(NM^{d-1}\right)\right)$, so the total cost will be between $\mathcal{O}\left(NM^{2(d-1)}\right)$ and $\mathcal{O}\left(N^2M^{3(d-1)}\right)$.

In comparison to the recursive Green's function approach we see that is is (asymptotically) worse for the asymptotically slowest way of computing the inverse and asymptotically better if we can use an algorithm that solves the sparse matrix-vector equation (4.2.0.1) in linear order. The rest of this section will be about solving the linear equation (4.2.0.1) using direct matrix methods, while in section 4.4 we will try to find an efficient iterative method for this problem.

4.2.1 LU-decomposition

We compare the performance of our implementation of the recursive Green's function algorithm to that of LU-decomposing the inverse Green's function G_{inv} (for a system with length N and slice size M^{d-1}) and then solving the set of equations $G_{inv}g_i = e_i$, where g_i denotes the i-th column of the Green's function, e_i the i-th standard basis vector in $\mathbb{R}^{NM^{d-1}}$ and the index i runs from $(N-1)M^{d-1}+1$ to NM^{d-1}.

The LU-decomposition is carried out by using the Matlab \-operator, which calls a build-in LU-decomposition algorithm. Since this is such a common operation, we may assume that it is implemented efficiently in Matlab and it should certainly be able to obtain the performance stated in appendix C and yield high accuracy results.

The bandwidth of G_{inv} is of the order of the slice size and the total dimension of the order of the system size, so if we plug those into the performance estimates of appendix C, we obtain an expected performance of $\mathcal{O}(NM^{3(d-1)})$, both for the LU-decomposition and the forward- and back-substitution [44]. This means the scaling of both the RGF and the LU based algorithms is asymptotically the same and we should use practical performance and accuracy tests to see which one is to be preferred.

4.2.1.1 Results

We tested the performance of both RGF and the built in LU-decomposition on a system of practical importance for our purposes: the tests where run on a Hamiltonian as described in section 5.2. This is a two-dimensional tight-binding Hamiltonian on a square lattice, with four *orbitals* per site.[5] The self-energy of this Hamiltonian has to be computed using the methods of section 3.4. We denote the number of sites in the direction along the leads as N_x and perpendicular to the leads as N_y, so we have $N = N_x$ and $M = 4N_y$.

We tested the execution time of both methods using the Matlab profiler. During these tests we fixed $N_y = 80$ and increased N_x. This particular test is of great practical importance, since computations like this need to be done many times to obtain the results in section 5.2. The computed answers of both methods coincided very well (and were sensible), so they are both stable in this case and the choice between them can be based on computational performance. The timings are summarised in table 4.2.

	LU	RGF
N_x	time (s)	time (s)
80	57	11
120	77	13
160	98	15
200	120	17
240	143	20

Table 4.2: Timings of the computational time for both LU-decomposition and RGF algorithms. Timings were done using Matlab R2012b on a Windows 7 Enterprise system using an Intel Core 2 Duo CPU E7400 @2.80GHz and 2GB of RAM.

From table 4.2 it can be seen that the increase of computational time does

[5] That is, there are four different electrons at each site: they are at the same position, but their full state (including for instance spin and chemical potential)is different, so the Pauli exclusion principle is not violated.

indeed seem linear when increasing N_x for both methods, but that the RGF algorithm performs much faster overall. It should be remarked that these systems are relatively small (so the asymptotic scaling may not tell the whole story at these sizes) and that for example background processes run on the system may cause inaccuracies in the timings. However, the difference in the time measurements are large and consistent and this test is realistic, so even this limited test provides good grounds to do further computations preferring the RGF algorithm over this LU-decomposition.

4.2.1.2 An alternative method for LU-decomposition

It is possible to speed up the LU-decomposition by reordering the matrix to be inverted. This effectively reduces the band width and hence the fill-in. The backsubstitution can be sped up by viewing the right hand side vectors as sparse and by exploiting the fact that only part of the matrix is needed. This is the approach used by [37]; they use the MUMPS library ([52] and references therein), in which all these functionalities have been integrated. A speed-up is reported for large systems. For practical reasons we however choose to not use this package, but just the basic Matlab functionality.

4.3 An analogous problem: the Helmholtz equation with radiation boundary conditions

As a motivation for the search for efficient iterative methods for finding the Green's function, we describe a related problem for which iterative methods have been shown to be effective. [53, 54, 55]

In section 3.3.4, we consider the transmission through a 2-dimensional uniform metal. If we go back to the time-independent Schrödinger operator (2.1.2.3) with zero vector potential, set $\frac{\hbar^2}{2m} = 1$ and define $k(\mathbf{x}) = \sqrt{-E - U(\mathbf{x})}$, we obtain the Helmholtz operator $\hat{A}_h(u, k) := -\Delta u - k^2 u$, where u is a function and $k(\mathbf{x})$ the wave number, which can in principle depend on the space coordinates.

In the above mentioned literature, [53, 54], the interest in the Helmholtz operator is primarily in the context of geophysical seismic imaging and the boundary conditions are chosen to be suitable for that context. This means that the shape of the domain is not necessarily rectangular and at the boundary a combination of Dirichlet and *first-order Sommerfeld radiation* conditions is specified. The first-order Sommerfeld radiation boundary condition is defined by

$$\frac{\partial u}{\partial \mathbf{n}} - iku = 0, \tag{4.3.0.1}$$

where **n** is the outward normal vector.

Applying a Sommerfeld radiation condition is of course different from attaching an ideal lead, but they do share some important features: it can be seen that a plane wave with wave vector k moving normal to the surface satisfies the radiation condition, while in an ideal lead modes will asymptotically be plane waves. Furthermore, in both cases the matrix to be inverted is sparse, complex symmetric (and not Hermitian) and indefinite [53, 54]. This makes it useful to consider methods used to solve the geophysical problem and see if they may be applicable to some transport problems.

4.3.1 Preconditioners

The approach used in [53, 54, 55] is to search for a good preconditioner and then apply a Krylov method (such as the generalized conjugate gradient method, used in [53] or GMRES, used in [55], see appendix C). Here we discuss briefly the algorithm of [53, 55], which uses a shifted Laplace preconditioner[6] combined with a projection procedure known as deflation which is used to get rid of near-zero eigenvalues that slow down the Krylov convergence. To define the projection a procedure based on multigrid method interpolations is applied [53, 55].[7]

To give a better idea of the method, we describe one variant of the algorithms given in [53, 55]. In this algorithm, a typical choice of preconditioning parameter c would be $1 + 0.5i$ and the operator $\hat{L_c}$ would not be inverted explicitly. In fact, the implicit inversion used during the (flexible) Krylov method would consist of only one multigrid iteration[8] [55], which means we do not actually precondition with $\hat{L_c}$, but just use it to find good search vectors [55, 56].

After one or more Krylov iterations we project the near-zero eigenvalues away; in our variant of the algorithm they will be projected to the largest eigenvalue, which is one[53].[9] After the projection procedure, a smaller system (the dimensions of the matrices are typically halved) results. With this procedure, the small system will have no (near)-zero eigenvalues and retains the structure of the original problem. This implies we can use the same

[6]This means we multiply both sides of the problem with M^{-1}, where M is the discretised version of the operator $\hat{L_c}(u, k) := -\Delta u - ck^2 u$, with a conveniently chosen $c \in \mathbb{C}$, see also appendix C

[7]Multigrid methods work by smoothening residuals and then mapping them to coarser grids. After having solved the problem on the coarsest grid, interpolations are used to obtain solutions on finer grids. From this interpolation operator a projection operator is defined in [53, 55].

[8]This means we apply a V-cycle, in which we go to the coarsest level once and back.

[9]See [55, 53] for more details and an explicit formula for this operator.

procedure again.

At some point we will arrive at a very coarse grid , for which we can solve the equation exactly. We then use interpolations to go back to finer grids, applying multigrid-preconditioned Krylov iterations again at each finer level. At the finest level an outer Krylov method is applied with as many iterations as needed to obtain an acceptable solution.

In summary, our algorithm will consist of applying one or more preconditioned Krylov steps, in which the preconditioner equations are approximately solved by a multigrid V-cycle and then a projection, at which point we can repeat the preconditioned Krylov iterations and a new projection. When we arrive at a very coarse grid the equation can be solved exactly. We then interpolate and refine our solution. At the finest level we then solve the system with a flexible Krylov method.[55]

This method does converge quickly; although the convergence speed does still depend on the number of grid points [55, 53], this dependence is not too strong. Though an implementation of a method comparable to this algorithm should certainly be less computationally intensive then either the RGF or some optimised LU method, especially in 3D, it should be clear from this discussion that it would be very complicated to implement and that it should be tuned separately for different problems, which makes it infeasible for us to implement.

4.4 Iterative methods for the Green's matrix columns

Considering the results of the previous section and noting the analogies to the system discussed in section 3.3.4, we test the performance of iterative methods to compute the matrix columns of the Green's function. In particular, we test the scaling of the number of iterations of the generalised minimal residuals method (see appendix C) as we increase the system size of the 2D uniform hopping model. In order to reduce the number of iterations, various preconditioners will be used.

The goal of preconditioning a matrix A with preconditioner M (where M is a non-singular matrix) is to obtain a modified matrix $M^{-1}A$ for which less GMRES-iterations are needed to converge to an accurate solution of $M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$.[10] Our test system is the uniform 2D hopping model described in section 3.3.4. As a benchmark and as a check of the GMRES

[10]We remark that we do in general not compute the explicit inverse M^{-1} and that we sometimes also precondition in a slightly more general way, i.e. we take matrices M_1, M_2 and solve $M_1^{-1}AM_2^{-1}M_2\mathbf{x} = M_1^{-1}\mathbf{b}$.

implementation[11] we first test $M = \hat{\mathbf{1}}$ (i.e. no preconditioning) and $M = A$ (i.e. using a direct method, so the number of GMRES iterations has to be 1). The result is shown in table 4.3.

Preconditioner	A	$\hat{\mathbf{1}}$
n	Iterations	Iterations
10	1	80
20	1	279
30	1	617
40	1	1080
50	1	1676

Table 4.3: Number of GMRES iterations without restarts needed to obtain a relative residual $< 10^{-6}$, for different preconditioners. The parameters $t = 1$, $\mu = 0$ and $E = 2$ are chosen; we always choose square systems with dimensions $n \times n$ (so the matrix size is $n^2 \times n^2$).

4.4.1 Preconditioning

Having tested these trivial preconditioners (see table 4.3) we conclude that preconditioning is indeed necessary for this problem and search for more sophisticated ways of preconditioning the system. We try to find good preconditioners using three different strategies: the first is to find a simplified version $\tilde{A}$ of the matrix A, for which it should be easier to solve equations (specifically, we can choose $\tilde{A}$ to be the corresponding Dirichlet problem or we can shift the energy to make $\tilde{A}$ positive definite; to solve the positive definite problems arising in the iteration steps we could for example use a multigrid method).

The second approach consists of preconditioning with an approximate inverse of A which is easy to compute (for example using an incomplete LU decomposition).

The third strategy is based on the realisation that we will often be interested in computing the transmission as a function of some parameter (energy, disorder, etc.); this means we want to invert parts of matrices that are all closely related. If we pick a representative matrix $\bar{A}$ and compute the (approximate or full) LU decomposition $\bar{L}$, $\bar{U}$, this may be a good preconditioner for a whole class of matrices with comparable parameters. A key observation here is that we need to compute the decomposition only once, so it does not matter if it is relatively slow.

[11] We use the inbuilt gmres()-routine in Matlab [46].

When implementing the first strategy mentioned above, we realize that shifting the energy to make the matrix positive (or negative) definite is a special case of the third strategy (with a relatively large shift). Because of this we first focus on preconditioning with the corresponding Dirichlet problem. The results are summarised in the first two columns of table 4.4. In some cases the preconditioner was ill-conditioned, causing the GMRES-algorithm to break down. If we look at the cases where preconditioning succeeded, we see that the number of iterations scales as $n^{0.8}$. This is better than the intuitive estimate which would be that the number of iterations should scale with the number of boundary sites, i.e. as n^1. It is however worse than the scaling $n^{0.5}$ required to make the method scale as well for this system as the RGF algorithm.

Preconditioner	Dirichlet	ILU(0)
n	Iterations	Iterations
10	13	41
20	-	78
30	25	5
40	32	5
50	-	4
60	36	4
70	40	4
100	55	3
150	66	3
200	-	3
250	95	3

Table 4.4: Number of GMRES iterations without restarts needed to obtain a relative residual $< 10^{-6}$, for different preconditioners. The parameters $t = 1$, $\mu = 0$ and $E = 2$ are chosen; we always choose square systems with dimensions $n \times n$ (so the matrix size is $n^2 \times n^2$). If the method does not converge, we write -.

In the third column of table 4.4, we show the number of iterations needed if we precondition the system with an incomplete LU decomposition with zero fill-in. That is, we approximate A with the product LU of a unit lower triangular matrix L and an upper triangular matrix U such that if there is an (off-diagonal) zero entry in A, there is also a zero entry in L and U. This means that LU will not be the real LU decomposition of A (since we disallow fill-in), but it approximates it.[50]

This ILU(0) approximation is the simplest form of the incomplete LU de-

composition. It can be computed extremely quickly and as can be seen in table 4.4, it is a very good preconditioner. For the problem considered in this section, the analytic computation of the self energy using 3.3.4.7 takes up much more computational time than the actual transport computation, but a comparison of executing times of either the RGF main loop or the GMRES-routine using the Matlab profiler does demonstrate GMRES to be almost twice as fast for the 250×250 system.

In order to test to what extent we can use preconditioners that differ from the matrix we want to partially invert, we take up the same system we study in the rest of the section and precondition with matrices with shifted energies. Both complete LU decomposition and ILU(0) decomposition are tested. The results are in table 4.5.

Interestingly, the preconditioner obtained with ILU(0) decomposition is much more robust to shifts in the energy, up to a certain point (which in this case would be around $\Delta E \approx 0.8$), after which the approximation is as bad as the full inverse. In any case the full inversion does not seem to be worth the effort, since it is very sensitive to energy shifts. The ILU(0) preconditioner may be used for systems that differ only a little, but since it can be computed very fast anyhow, it is doubtful whether it would help speed up the computations.

Preconditioner	$A(E + \Delta E)$	ILU(0) of $A(E + \Delta E)$
Shift ΔE	Iterations	Iterations
0	1	5
0.05	23	5
0.1	39	6
0.2	71	6
0.5	160	6
1	253	298
2	337	343

Table 4.5: Number of GMRES iterations without restarts needed to obtain a relative residual $< 10^{-6}$, for different preconditioners. The parameters $t = 1$, $\mu = 0$, $E = 2$ are chosen; we always choose square systems with dimensions 30×30 (so the matrix size is 900×900).

In summary, there are many different ways of computing the transmission function and some perform much better than others. For the 2D test problems in this chapter, the recursive Green's function algorithm has done reasonably well, but it's not as fast as highly optimised direct solvers or Krylov subspace methods with the right preconditioner. If we wanted to compute

the transmission through a large 3D system, it is expected that the difference between those methods would become larger and that the iterative methods would be the fastest among them.

We do however intend to use the recursive Green's function algorithm for the systems considered in the following chapter, because we already have a working algorithm available in the commercially available software Matlab. To implement a more efficient method (especially in the case of the most efficient but parameter dependent preconditioned iterative methods) would require a significant investment of "human time" in terms of the analysis, programming and optimisation necessary. It is furthermore likely that slightly different methods or parameters should be used for different systems. It is therefore more practical to invest computer time instead of human time and keep using the efficient but slightly suboptimal RGF algorithm if possible.

Chapter 5

Topological insulators

With the methods described in previous chapters, we can compute conductances through physical systems of interest. Here we focus on *topological insulators*. These are systems with the band structure of insulators that do however have non-zero conductance if non-periodic boundary conditions are applied. This conductance can be attributed to modes lying close to the boundary of this system, but the exisistence of these boundary modes can be determined by considering the symmetry properties of the bulk (i.e. periodic) system [1]. Another remarkable property is that disorder does not hinder the conductance of these boundary modes, as long as the symmetries of the system are respected [1].

We first discuss some theory of topological insulators and demonstrate important properties using an example. We then turn to a non-standard topological insulator, for which one of the protecting symmetries lies in the lattice it is defined on. For this system it is known (and will be demonstrated) that boundary modes do exist, but their stability under disorder had not yet been tested. This section 5.2 is based on [9].

5.1 Topological states of matter

The standard theory of topological states of matter is based on classifying Hamiltonians according their anti-unitary symmetries. These symmetries are particle-hole symmetry and time-reversal symmetry. By making the assumption that other (unitary) symmetries play no role, a theory for which disorder plays no role can be derived. It turns out that we can distinguish ten different classes of topological insulators and superconductors this way (in any dimension) [1, 2, 3]. In this thesis our focus will be on 2D systems with time-reversal symmetry.

According to the above mentioned ten-fold classification, to systems with

certain symmetries and dimensionalities we can assign topological indices, which are homotopy invariants of the band structures the (periodic) Hamiltonians encode. We would like to study the effect of the boundary conditions on the band structure. It turns out that for insulators these effects can be described by looking at the topological indices [1, 2, 3, 4]. Depending on these indices, we can distinguish two possible dependencies on the boundary conditions: if the spectrum of a system where the periodic boundary conditions are replaced by zero Dirichlet conditions is a projection of the band structure along the axis where the system is "cut open", we call the insulator trivial. A topological insulator is a system where replacing the periodic boundary condition by Dirichlet boundary conditions give rise to extra (boundary) modes in between the bands.

5.1.1 Bloch's theorem revisited

Before we give some examples of topological insulators and describe some of their properties, we will first develop a theoretical framework to interpret them in. This will be based on considering the results from section 2.1.2 in a more abstract way. These considerations are based on [57, 20, 1].

Consider a system with Brillouin zone B and $n + m$ bands and assume without loss of generality that the chemical potential is at 0, there are n bands below zero over the whole of B and m bands above. Picking as a basis the different bands, the energy functional $E[\Psi] := \langle\Psi| \hat{H} |\Psi\rangle$ becomes a function from $B \times \mathbb{C}^{m+n}$ to $\mathbb{R}$.

For any $\mathbf{k} \in B$, we can write an orthonormal basis with $n + m$ eigenstates. This means that choosing a different $\mathbf{k}' \in B$ amounts to choosing a different basis for $\mathbb{C}^{m+n}$. In general, switching between two bases of $\mathbb{C}^{m+n}$ is encoded by picking an element of the group $U(m + n)$.

However, there are two special states: the ground state, for which all bands with negative energies are occupied and the highest-energy state, for which all bands with positive energies are occupied. These states are the same over the whole Brillouin zone and are stabilised under a group action of $U(n)$ and $U(m)$ respectively. Because of this, the different choices of momentum will only correspond to the action of a subgroup of $U(m+n)$ and we can interpret the Hamiltonian as a map $H : B \to G_{m+n,n} := U(m + n)/(U(m) \times U(n))$.

Viewing the Hamiltonian as a function from the Brillouin zone to some Grassmann manifold[1] of which the properties only depend on the number of occupied and empty bands gives us the possibility to classify Hamil-

[1] See [58]

tonians using homotopy theory. If we for example take B to be the 2-torus[2] $\mathbb{T}^2$ and consider insulators with 1 filled and 1 empty band, then $H : B = \mathbb{T}^2 \to G_{m+n,n} \simeq S^2$. The winding number of these maps is a homotopy invariant [59], [60] and can be any integer. This winding number $C_1 \in \mathbb{Z}$ is also called the *first Chern number*.

5.1.2 Chern numbers and Dirac cones

As an example of an insulator with a topological structure, we consider a *Chern insulator*[61]. The model we consider is slightly different from the original and is described in [20]. The Hamiltonian will in this case be a continuous function $H : \mathbb{T}^2 \to S^2$.[3]

We can write these Hamiltonians as 2×2 matrices, for which it is convenient to decompose them in terms of a basis consisting of the 2×2 identity matrix $\mathbf{1}$ and the three Pauli-matrices, which are given by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The Hamiltonian can now be written as

$$H(\mathbf{k}) := d_0(\mathbf{k})\mathbf{1} + \sum_{i=1}^{3} d_i(\mathbf{k})\sigma_i =: d_0(\mathbf{k})\mathbf{1} + \mathbf{d}(\mathbf{k}) \cdot \sigma, \tag{5.1.2.1}$$

where $d_0(\mathbf{k})...d_3(\mathbf{k})$ are scalar functions defining the particular Hamiltonian[4] and introduced the vectors $\mathbf{d}(\mathbf{k}) := (d_1(\mathbf{k})\, d_2(\mathbf{k})\, d_3(\mathbf{k}))^T$ and $\sigma := (\sigma_1\, \sigma_2\, \sigma_3)^T$ for brevity.

If we now take $d_0(\mathbf{k}) = 0$ (this has no effect on the topological properties of the Hamiltonian), assume $||\mathbf{d}(\mathbf{k})|| \neq 0$ (for all $\mathbf{k}$) and define $\hat{\mathbf{d}}(\mathbf{k}) := \frac{\mathbf{d}(\mathbf{k})}{||\mathbf{d}(\mathbf{k})||}$, the winding number can be computed [20]

$$C_1 = \frac{1}{4\pi} \int_0^{2\pi} \mathrm{d}k_1 \int_0^{2\pi} \mathrm{d}k_2\, \hat{\mathbf{d}}(\mathbf{k}) \cdot \left(\frac{\partial \hat{\mathbf{d}}(\mathbf{k})}{\partial k_1} \times \frac{\partial \hat{\mathbf{d}}(\mathbf{k})}{\partial k_2} \right). \tag{5.1.2.2}$$

We can define Chern Hamiltonians $H_r(\mathbf{k})$ by $(d_1(\mathbf{k})\, d_2(\mathbf{k})\, d_3(\mathbf{k}))^T = (\sin(k_1)\ \sin(k_2)\, r + cos(k_1) + cos(k_2))^T$, where r is a real parameter. We remark that for some

[2] This particular Brillioun zone arises if we take a 2D rectangular system and apply periodic boundary conditions in both directions. It is a very common geometry to consider in physics, especially if the radii of the torus are taken to be large.

[3] To be more precise, the Hamiltonian will be a continuous function with domain homomorphic to $\mathbb{T}^2$ and co-domain homomorphic to S^2

[4] and for which some restrictions hold, since the codomain of the physical Hamiltonian should be (homomorphic to) S^2

combinations of $(r, \mathbf{k})$, the assumption $||\mathbf{d}(\mathbf{k})|| \neq 0$ does not hold any more.[5] For these values of r, the system is not an insulator. In fact, there are phase transitions taking place at $r = -2, 0, 2$, in which $C_1(r)$ changes. Specifically, we have [20]

$$C_1(r) = \begin{cases} 0 & \text{if } r < -2, \\ 1 & \text{if } -2 < r < 0, \\ -1 & \text{if } 0 < r < 2, \\ 0 & \text{if } r > 2. \end{cases}$$

The values for $\mathbf{k}$ at which there are closures of the band gap correspond to special points (called *high-symmetry points*) in the Brillioun zone. For later reference, we name the point $\mathbf{k} = (0, 0)$ the Γ-point, $\mathbf{k} = (\pi, \pi)$ the M-point, $\mathbf{k} = (\pi, 0)$ the X-point and $\mathbf{k} = (0, \pi)$ the Y-point.

If we would look at the band structure of the Chern Hamiltonians around these special points, we can use linear approximations to the Hamiltonians. For example, around $(r, k1, k2) = (0, \pi, 0)$, the Hamiltonians can be approximated by $H_r(\mathbf{k})|_{(0,\pi,0)} \approx k_1\sigma_1 + (\pi - k_2)\sigma_2 + r\sigma_3$, which is a two-dimensional Dirac Hamiltonian (with momentum $(k_1, \pi - k_2)$ and mass r).[6]

Around the other special points the Chern Hamiltonians can also be approximated well by Dirac Hamiltonians. The regions where these approximations hold are called *Dirac cones*; they are the only places where the band gap closes. The jumps in the value of the first Chern number can be computed entirely from the Dirac approximations, but the actual value can only be determined by the full Hamiltonian.[20]

5.1.2.1 Bulk and boundary properties

Hamiltonians defined on a torus are known as bulk Hamiltonians. The reason for this nomenclature is that the boundary conditions are chosen such that the analysis of the Hamiltonian is easy (specifically, Fourier transforms are possible), but without regarding the physical situation where most systems do not live on tori. The assumption is that for properties that depend on how states behave in the middle of the system (the bulk), the precise boundary conditions do not matter and the bulk Hamiltonian is accurate.

We can however also choose to pick different boundary conditions and in the case of topological insulators it turns out to also be interesting to study these Hamiltonians on cylinders (with Dirichlet boundary conditions on the edge) and in the lead-central region-lead geometry discussed in the previ-

[5] This is the case for $(r, k1, k2) = (-2, 0, 0)$ or $(0, 0, \pi)$ or $(0, \pi, 0)$ or $(2, \pi, \pi)$

[6] In $3+1$ dimensions, the Dirac equation describes the quantum mechanical behaviour of relativistic electrons and other relativistic spin-$\frac{1}{2}$ particles.

ous chapters. We will not do any transmission calculations for the Chern insulator, but we will discuss boundary effects arising from considering the system on a cylinder.

The procedure for placing the system on a cylinder consists of performing a (discrete) Fourier transform of one of the directions (for example transforming the wave vector k_1 to the lattice coordinate x) and setting Dirichlet boundary conditions in this transformed direction, after which the spectrum can be obtained numerically for every value of k_2. If we do this for many values, we can plot the spectrum on a vertical line for each value of k_2. If we then connect the largest values of the spectra for each k_2 to each other, the second largest to each other etcetera, we get one line for every site in the x-direction. Together these lines form a band structure, of which an example is plotted in figure 5.1.[7]

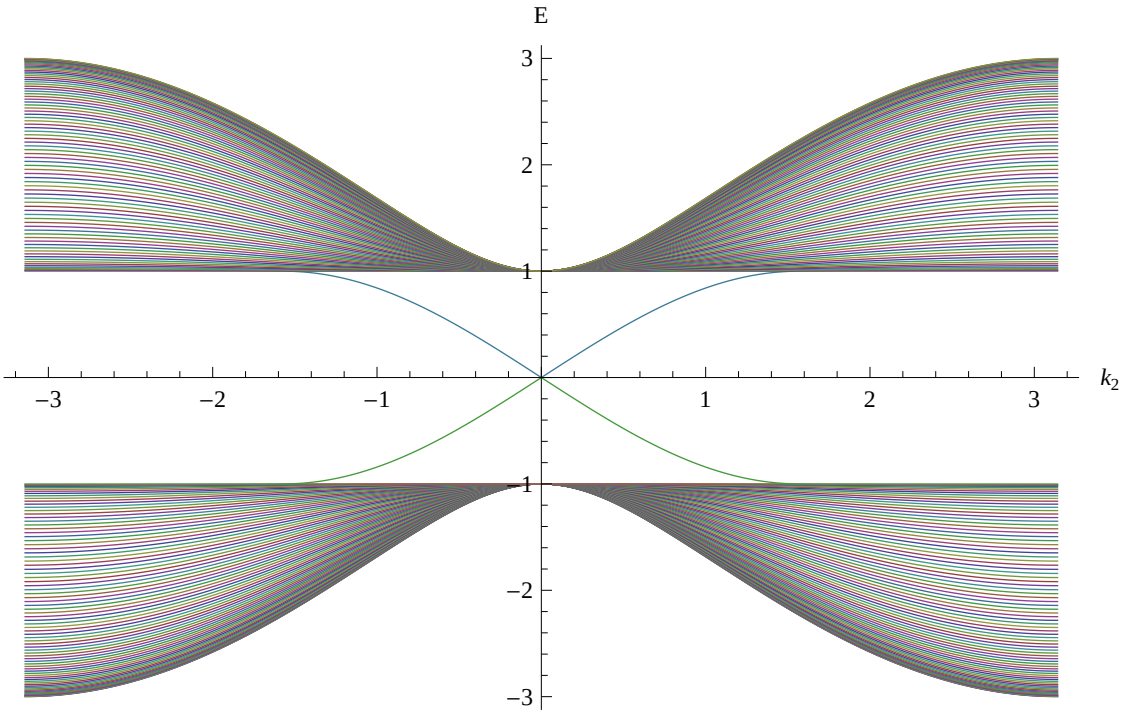

Figure 5.1: The band structure of a Chern insulator with $r = -1$ and 80 sites in the x-direction. k_2 was discretised with step size $\frac{\pi}{100}$.

[7]Please note that with this procedure the modes are plotted as if they never cross. This is a numerical artefact, since in reality modes can and do cross. For the study of boundary conditions, this simplified picture of the bulk band structure is however sufficient, since we only need to know at which energies the bulk lies and (if they exist) how the boundary modes look like.

Most of figure 5.1 looks the same as how a projection of the bulk spectrum would look. The main differences are that the two bands now consist of many individual lines (an unimportant finite size effect) and that there are two modes in between the bands that cross at the point k_2 is zero (an important consequence of applying different boundary conditions). These two modes are called *boundary modes*.

The name boundary mode is chosen appropriately for multiple reasons: even though there existence can be argued for on the basis of bulk properties alone (the topological quantum numbers), they only manifest themselves in a geometry with boundaries (such as the cylinder) and their position is also at the boundary (i.e. the probability density function of the position is strongly concentrated close to the boundary).

Further properties of the two boundary modes are that they each are located at a different boundary with opposing group velocities. The corresponding wave functions are perpendicular to each other. This perpendicularity has important consequences for the stability of the boundary modes under disorder.

The effect of the chemical potential disorder we have considered, is that it may randomise the momentum of an electron while keeping the energy constant.[8] The electron will be scattered into another mode, with a probability of this event occurring proportional to the disorder strength and the overlap integral between the modes.

In the examples considered in section 3.3 or in the bulk of the Chern insulator, at energies where there is transmission, there are many modes available with different group velocities and non-zero overlap. This implies that the disorder can induce events which scatter electrons into other (counter-propagating) modes and thus hinder transmission.

The situation is different if we are in the bulk gap of the Chern system: for a given energy in the gap, the only scattering events that comply with the conservation of energy scatter an electron from one edge mode into the same edge mode (in which case nothing happens and transmission is not hindered) or into the other edge mode, but this last event has a probability of almost zero, since the overlap integral between edge modes is exponentially suppressed as a function of system width. The conclusion is that the transmission is essentially unhindered by disorder, making this system very robust. This key property is one of the main reasons physicists study topo-

[8] Disorder also changes the spectrum, but we assume it is small enough not to effect it qualitatively.

logical insulators.

Because the fact that boundary modes for which the group velocities are in different directions do not have any overlap is so important for the way these systems behave under disorder, we will demonstrate this property in the context of a Chern insulator. To do this we linearise the edge state theory around $k_2 = 0$. Around this point the dispersion is linear and the two edge modes have velocities with the same magnitude v but different directions; this can be expressed with an effective 1D massless Dirac equation[9]

$$\hat{H}_{eff}(k_2)\Psi := \begin{pmatrix} 0 & \hbar v k_2 \\ \hbar v k_2 & 0 \end{pmatrix} \Psi = E\Psi. \tag{5.1.2.3}$$

We should now consider solutions to (5.1.2.3) with the same energy but different momenta.

The eigenvectors solving (5.1.2.3) are $\Psi_{\pm}(k) = \frac{1}{2}\begin{pmatrix} \sqrt{2} \\ \pm\sqrt{2} \end{pmatrix}$ and they correspond to (orthogonal) solutions with the same momentum k and different energies $\pm\hbar v k$. This means that if we define $\Psi^{\pm}(k) := \Psi_{\pm}(\pm k)$, we have found the solutions with the same energy and different momenta. Since $\Psi_{\pm}(k)$ does not depend on k, it is obvious that the same energy solutions $\Psi^{\pm}(k)$ are also orthogonal.

There is a simple way to destroy the orthogonality: if we add a small magnetic field, the Hamiltonian in (5.1.2.3) is modified by the addition of an extra term $B\sigma_3$, where B is the magnetic field strength, assumed to be small, so as not to invalidate our effective description.

We get $\hat{H}_{mag}(k_2, B) = \hat{H}_{eff}(k_2) + B\sigma_3$, with perturbed energies $E((k_2, B)) = \pm\sqrt{B^2 + (\hbar v k_2)^2}$ and eigenvectors $\Psi_{\pm}(k_2, B) \sim \begin{pmatrix} \frac{B \pm E(k_2, B)}{\hbar v k_2} \\ 1 \end{pmatrix}$. The energy remains an even function of the momentum and we should compute the overlap integral $\langle \Psi_+(k_2, B) | \Psi_+(-k_2, B) \rangle = -\frac{2B}{(\hbar v k_2)^2}(B + E)$. If we send $B \to 0$ we re-obtain the results for the system without magnetic field, but for non-zero magnetic field the boundary spectrum becomes gapped and the remnants of the boundary modes start to overlap, ending the robustness against disorder.

We remark that at the band gap closures the Chern Hamiltonians commute with the time-reversal operator. This means that the time-reverses of eigenstates of the Hamiltonian must also be eigenstates and that if these

[9] The dispersion relation $E(k_2) = \pm\hbar v |k_2|$ is immediate from (5.1.2.3)

eigenstates are different they are orthogonal. Since Kramer's degeneracy theorem [62] states these eigenstates must be different, eigenstates come in pairs that are each others time reversal invariants. The boundary modes we encountered are such a pair. Because an external magnetic field breaks the time-reversal invariance of the Hamiltonian, this restriction no longer applies and the orthogonality can be broken.

5.2 Topological crystalline insulators

We have discussed some general properties of topological insulators, using the Chern insulator as an example. In these systems, certain symmetries can give rise to topological invariants and systems with nontrivial topological invariants can have special boundary modes. However, the symmetries introduced before are not all symmetries one can think of. We could for example also look at whether the symmetries of the underlying crystal lattice gives rise to comparable boundary modes. It turns out there are indeed boundary modes [4, 5, 6, 7, 8].

These boundary modes can also arise in systems that are trivial according to the tenfold way [4, 5, 6, 7, 8]. The effect of disorder on these systems remained however unclear [9]. In this section we will address this question for one concrete example of such a system; it is based on a paper [9], published together with Carolin Küppersbusch, Vladimir Juričić and Lars Fritz.

The starting point is a model proposed in [5], which is an extension of the Bernevig-Hughes-Zhang model [63]. The Hamiltonian we consider is defined on a 2D square lattice and its Fourier space representation is given by [5, 9]

$$\mathcal{H} := \sum_{\mathbf{k}} \Psi_{\mathbf{k}}^{\dagger} \begin{pmatrix} H(\mathbf{k}) & H_{\mathrm{SO}}(\mathbf{k}) \\ H_{\mathrm{SO}}^{\dagger}(\mathbf{k}) & H^{*}(-\mathbf{k}) \end{pmatrix} \Psi_{\mathbf{k}}, \tag{5.2.0.4}$$

where the four components of Ψ are two different orbitals (p and s) with two different spins ($\uparrow$ and $\downarrow$); when the Hamiltonian is written as in (5.2.0.4), these orbitals are ordered as $\left(s_{\mathbf{k}}^{\uparrow}, p_{\mathbf{k}}^{\uparrow}, s_{\mathbf{k}}^{\downarrow}, p_{\mathbf{k}}^{\downarrow}\right)$. Here $H(\mathbf{k})$ is given as the dot product of a vector of Pauli matrices σ and a vector $\mathbf{d}(\mathbf{k})$ (which also depends on the model parameters M, B and $\tilde{B}$)

$$H(\mathbf{k}) := \mathbf{d}(\mathbf{k}) \cdot \sigma \quad \text{with} \quad \mathbf{d}(\mathbf{k}) := \begin{pmatrix} \sin k_x + \cos k_x \sin k_y \\ -\sin k_y + \cos k_y \sin k_x \\ M - 2B\left[2 - \cos k_x - \cos k_y\right] - 4\tilde{B}\left[1 - \cos k_x \cos k_y\right] \end{pmatrix}. \tag{5.2.0.5}$$

The spin-orbit coupling Hamiltonian can be used to couple the spin-up sector to the spin-down sector. We define this *Rashba* coupling term (with coupling

strength R_0) as [5][10], [9]

$$H_{\mathrm{SO}}(\mathbf{k}) := R_0 \begin{pmatrix} -i\sin(k_x) - sin(k_y) & 0 \\ 0 & 0 \end{pmatrix}. \tag{5.2.0.6}$$

We first analyse the phase diagram of the model for $R_0 = 0$. In this case the system admits four phases [5].[11] The boundaries between the phases can be found by searching for parameters such that $\mathbf{d}(\mathbf{k}) = 0$, for some values of $\mathbf{k}$. This happens along the line $\frac{M}{B} = 8$ (with $\mathbf{d}(\pi, \pi) = 0$) and along the line $\frac{M}{B} = 4 + 8\frac{\tilde{B}}{B}$ (with $\mathbf{d}(0, \pi) = 0$ and $\mathbf{d}(\pi, 0) = 0$).[12] We note that to go from the trivial state with $M < 0$ to the trivial state wtih $\frac{M}{B} > 4 + 8\frac{\tilde{B}}{B}$ and $\frac{M}{B} > 8$, we will always pass through phase transitions such that at each high-symmetry point (Γ-, X-, Y- and M-point) the bulk gap will close once.

The next step is identifying what these four phases are. This has been done in [5] in terms of symmetries of the system. These symmetries are time-reversal symmetry (so the Chern number of the complete system is zero, but for zero spin-orbit coupling we can assign Chern numbers to the different spin sectors) and the underlying lattice symmetry (for which the high-symmetry points are important). In [5], the phases are named after the high-symmetry point at which the band gap closes in the transition to a trivial insulator. We follow the convention of [9] and refer to the Γ- and M-phases as *quantum spin hall insulator* (QSHI) and to the XY-phase as *topological crystalline insulator* (TCI). It is our eventual goal to compare the stability of the edge modes under disorder in both the QSHI and the TCI phase, with and without Rashba coupling. In order to do this it is sufficient to set $\tilde{B} = B$. Furthermore, B only functions as a scaling parameter, so it makes sense to consider values of variables in units of B; hence we define $m := \frac{M}{B}$ and $r_0 := \frac{R_0}{B}$.

For $r_0 = 0$, we can consider the subsystems consisting of only spin up or only spin down particles and compute[13] their Chern numbers $C_1^{\uparrow}$, $C_1^{\downarrow}$ using

[10] This definition differs from the coupling in eq. (S9) of [5] by an (unimportant) factor $\frac{1}{2}$.

[11] We make the (physically sensible) assumptions that $0 \leq \tilde{B} \leq B$.

[12] At $M = 0$ there is also a phase boundary ($\mathbf{d}(0, 0) = 0$).

[13] This computation was carried out with the help of Wolfram Mathematica; the integral was computed numerically for a value of m in each phase. Because Chern numbers only change at band closures and we know when the bands close, this is sufficient to determine them.

(5.1.2.2)

$$
\begin{aligned}
C_1^{\uparrow} &= \frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}k_1 \int_0^{2\pi} \mathrm{d}k_2\, \hat{\mathbf{d}}(\mathbf{k}) \cdot \left(\frac{\partial \hat{\mathbf{d}}(\mathbf{k})}{\partial k_1} \times \frac{\partial \hat{\mathbf{d}}(\mathbf{k})}{\partial k_2} \right) \\
&= \frac{\sqrt{2}}{4\pi}\int_0^{2\pi} \mathrm{d}k_1 \int_0^{2\pi} \mathrm{d}k_2 I \\
&= \begin{cases} 0 & \text{if } m < 0, \\ -1 & \text{if } 0 < m < 8, \\ -2 & \text{if } 8 < m < 12, \\ 0 & \text{if } m > 12, \end{cases}
\end{aligned} \tag{5.2.0.7}
$$

where

$$
\begin{aligned}
I = \Big[& \big(-(m-12)\sin(2k_1)\sin(k_2) + (m-12)\sin(k_1)\sin(2k_2) - 2\cos(k_1)((m-4)\cos(k_2) + \cos(2k_2) + 3) \\
& + \cos(2k_1)(-2\cos(k_2) + 2\cos(2k_2) - m + 6) - m\cos(2k_2) - 6\cos(k_2) + 6\cos(2k_2) - 6\big) \\
& \Big/ \big(8\cos(kx)(2\cos(ky)+1)(2\cos(ky)+m-8) + 2\sin(2kx)\sin(ky) - 2\sin(kx)\sin(2ky) \\
& + 32\cos^2(kx)\cos(ky) + \cos(2kx)(7\cos(2ky)+11) + 8(m-8)\cos(ky) + 11\cos(2ky) + 2(m-16)m + 147\big)^{3/2} \Big]
\end{aligned}
$$

and

$$
C_1^{\downarrow} = -C_1^{\uparrow}. \tag{5.2.0.8}
$$

From this observation the phase diagram for the system *with* Rashba coupling can be obtained, as long as the coupling is too weak to close the bulk band gap. The phase diagram is shown in figure 5.2 [9]. It should however be noted that assigning a Chern number to a certain spin band is only possible if $r_0 = 0$, since the Rashba coupling breaks spin conservation.[14]

5.2.1 Cylindrical geometry

Having described the system in Fourier space (i.e. on a torus), we want to "open it up" and study it on (more realistic) geometries that actually have boundaries. The procedure consists of performing a Fourier transform to the Hamiltonian (5.2.0.4) and subsequently applying non-periodic boundary conditions (zero Dirichlet conditions in one direction and open boundary conditions in another). We start by Fourier transforming in one direction and applying the Dirichlet conditions, analogously to section 5.1.2.1.

The way to obtain the Hamiltonian on a cylinder or fully in real space is straightforward but tedious, so we just demonstrate this for one term.

[14] For $r_0 = 0$ the $\mathbb{Z}_2$ topological invariant of interest ν is given by $\nu = \frac{C_1^{\uparrow} - C_1^{\downarrow}}{2} \bmod 2$. For the weak Rashba coupling discussed above the invariant does not change and remains well defined. We refer to [5] and references therein.

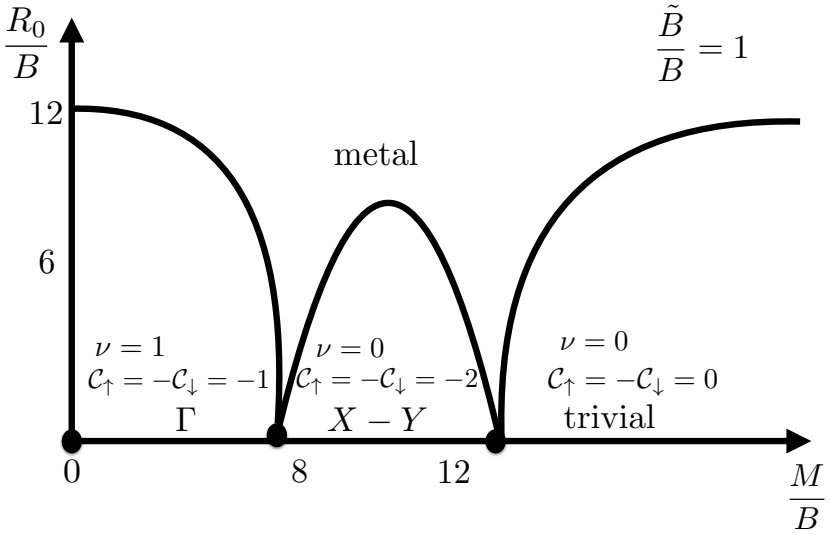

Figure 5.2: The phase diagram corresponding to (5.2.0.4), for certain parameter values. Figure taken from [9].

Suppose we want to transform in the x-direction, in which we assume our grid to have N_x sites. The vector $\mathbf{d}(\mathbf{k})$ contains the term $\cos(k_x)$, so the full Hamiltonian should contain a term $\sim \sum_{\mathbf{k}} \Psi^\dagger_{\mathbf{k}} \cos(k_x) \Psi_{\mathbf{k}}$. Then

$$\begin{aligned}
\sum_{\mathbf{k}} \Psi^\dagger_{\mathbf{k}} \cos(k_x) \Psi_{\mathbf{k}} &= \frac{1}{N_x} \sum_{\mathbf{k}} \sum_{x} \sum_{x'} \Psi^\dagger_{x,k_y} e^{-ik_x x} \cos(k_x) e^{ik_x x'} \Psi_{x',k_y} \\
&= \frac{1}{N_x} \sum_{\mathbf{k}} \sum_{x} \sum_{x'} \Psi^\dagger_{x,k_y} e^{-ik_x x} \frac{e^{ik_x} + e^{-ik_x}}{2} e^{-ik_x x'} \Psi_{x',k_y} \\
&= \frac{1}{2} \sum_{\mathbf{k}} \sum_{x} \sum_{x'} \Psi^\dagger_{x,k_y} \frac{e^{ik_x(x'+1-x)} + e^{-ik_x(x'-1-x)}}{N_x} \Psi_{x',k_y} \\
&= \frac{1}{2} \sum_{k_y} \sum_{x} \sum_{x'} \Psi^\dagger_{x,k_y} \left(\delta_{x'+1,x} + \delta_{x'-1,x} \right) \Psi_{x',k_y} \\
&= \frac{1}{2} \sum_{k_y} \sum_{x} \left(\Psi^\dagger_{x+1,k_y} \Psi_{x,k_y} + \Psi^\dagger_{x-1,k_y} \Psi_{x,k_y} \right).
\end{aligned}$$

This means that the term $\cos(k_x)$ corresponds to nearest neighbour hopping in the x-direction.[15] The zero Dirichlet boundary conditions can be applied straightforwardly, by not connecting the first and last site of the

[15] So following the analogy of section 2.1.3.1 we could interpret a term like $\cos(k_x) - 2$ as a standard approximation to a second derivative.[21]

grid. The term $\sin(k_x)$ corresponds to anisotropic nearest neighbour hopping $\sim (\Psi^\dagger_{x+1,k_y}\Psi_{x,k_y} - \Psi^\dagger_{x-1,k_y}\Psi_{x,k_y}$ (this is comparable to a leapfrog discretisation of a first derivative [21]).

Having obtained the Hamiltonian on a cylinder, we can get an idea of the band structure, analogously to what is done in section 5.1.2.1. In figure 5.3 example band structures for both the QSHI and the TCI phases are shown.

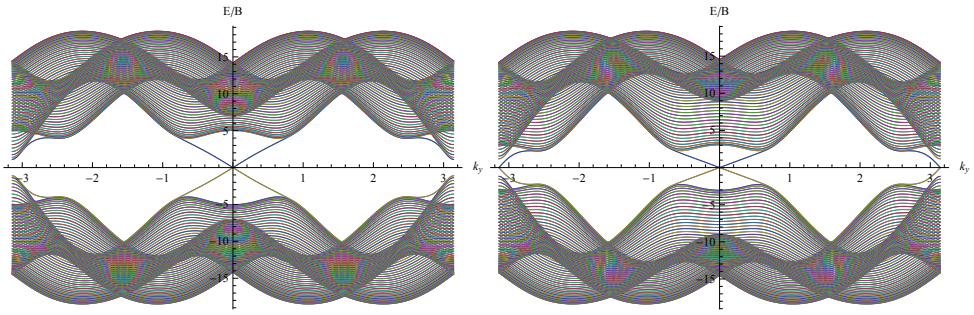

Figure 5.3: The band structure in the QSHI-phase (left, with $m = 7$, $r_0 = 0$ (and $\tilde{B} = B$)) and a TCI (right, with $m = 9$, $r_0 = 0$ (and $\tilde{B} = B$)). We took 80 sites in the x-direction and discretised k_y with step size $\frac{\pi}{100}$.

In figure 5.3 we see boundary modes that cross at $k_y = 0$ (both phases) and $k_y = \pi$ (TCI). We remark that if we project along the x-direction, the X- and Γ-points correspond to $k_y = 0$ and the Y-point to $k_y = \pi$. Taking into account that the spectrum is twice degenerate, we can simply count the boundary modes to conclude that if we connect the fully real space Hamiltonian to leads we expect a transmission $T(E)$ of 2 in the QSHI and 4 in the TCI.

We can now check numerically that the boundary modes are indeed located at the boundary of the system and how close to the boundary they actually remain. We can then estimate a suitable width of the system we shall use our transport algorithms on by requiring the overlap between modes at different boundaries to be (very) small. The reason we require this is that we want to study the properties of "large" systems.

In large systems, the boundaries of the system are properly separated from each other. For our purposes we can define the part of the system where a boundary mode has a non-vanishing[16] probability density to be the bound-

[16]We have to impose some cutoff here, but a precise definition is not given, since we expect the probability density to decay exponentially anyhow

ary (so it gets a certain thickness)[17] and the requirement that different boundaries do not overlap defines what it means to be a large system.

The test was performed by computing the eigenvectors Ψ at zero momentum (for some representative parameter values, such as the ones used in figure 5.2) and taking the four with the lowest energy (these are two right-going and two left-going modes). Then the probability density $|\Psi|^2$ was computed at every site. These probability densities can then be plotted to verify that the modes are indeed at the boundary. A second test consists of considering the sites that are in the middle and checking if the probability density is sufficiently small there. These checks show the modes to be indeed located at the boundary, but extending inwards somewhat. For a system of width 40 this inward extension is such that the boundaries (almost) touch each other, making this system size to small for our purposes. In a system of width 80 this is not a problem any more and the boundaries are really separated from each other.

Last but not least, we can use the properties of the system on a cylinder to come up with an idea of how the system would behave if we apply disorder. This is done by assuming the boundary modes to consist of the product of some complicated function $\Psi(x, s, o)$ depending on x, spin s and orbital o (which is the eigenvector of this mode in the cylindrical geometry) and a plane wave in the y-direction and furthermore assuming that disorder causes these plane waves to scatter elastically, such that energy is conserved and the probability of scattering into another mode $\tilde{\Psi}(x, s, o)$ with the same energy is proportional to disorder strength and the overlap integral $\left\langle \Psi | \tilde{\Psi} \right\rangle$.

Because we can numerically obtain the functions $\Psi(x, s, o)$ at some given energy in the bulk band gap and compute their overlaps, we can check if this overlap integral is zero[18] for all counter-propagating modes at that energy or not. For such a given energy close to zero we obtain two left- and two right-moving modes in each Dirac cone (so with momenta k_y either close to 0 or to π, depending on the phase). This means that in the QSHI we get two left- and two right-movers, while in the TCI we get four of both. When we compute the overlap between a left-mover and a right-mover at the same Dirac cone, we do see they are orthogonal (which they have to be because the system is time-reversal symmetric, so we get two Kramer's pairs [62]). This holds in both phases and with or without Rashba coupling.

In the TCI phase we should also check for overlap between left- and right-

[17] With this definition of the boundary it will in fact still consist of the physical boundary and (possibly) some neighbouring layers

[18] or rather extremely small, since our system is of finite size

movers at different Dirac cones. Here we do note a difference between the cases with and without Rashba coupling: if there is no Rashba coupling, there is no overlap between left- and right-movers at the different cones, while for $r_0 \neq 0$ the modes at different cones do start to overlap.

Of course, it is in general a rather special property for these functions to not overlap, so we should search for an explanation for this in the system without Rashba coupling. This has to do with the fact that the overlap also depends on the overlap between spin orientations and we have already seen that in the absence of Rashba coupling we can split the system up into two independent spin sectors, each with its own Chern number. Because modes in one of these sectors always have the same spin as other modes in that sector and always an opposite spin to modes in the other sector, the overlap integral between modes in different sectors has to vanish. Since counter-propagating modes will belong to different spin sectors, scattering between them is forbidden by spin conservation. Rashba coupling breaks this spin conservation and hence scattering between counter-propagating modes not related by time reversal symmetry becomes allowed again.

In summary, our analysis of this model on a torus and cylinder has enabled us to predict that we will find both QSHI and TCI phases, with bulk gap conduction $G = 2e^2/h$ in the QSHI phase and $G = 4e^2/h$ in the TCI phase. If we put disorder into the system we do not expect these conductances to change[19], unless we are in the TCI phase with non-zero Rashba coupling, in which case we do expect to see Anderson localisation.

5.2.2 Conductance and disorder

We now turn to the system in real space and want to connect it to leads. In order to come up with a realistic coupling, we let the system and leads differ only in chemical potential and Rashba coupling (in the leads we pick μ in the bulk band and $r_0 = 0$). The lead self-energy is then computed using the methods described in section 3.4.[20] We remark that strictly speaking this process already breaks the rotational symmetry of the crystal lattice, but we expect this to be a minor effect.

Having computed the self-energy, we can use the methods of chapter 4 to obtain the conductance. As described in section 4.2.1.1, we can do this both with the RGF and the LU algorithms. The results are the same (and sensible), indicating both methods to be stable, but the RGF algorithm is much faster. The results in the rest of this section are all obtained using the RGF

[19] As long as the disorder is too weak to close the bulk gap.

[20] These methods turn out to work only for some values of μ, but we can in each phase still pick a suitable value of μ that is properly in the band.

algorithm.

We want to compare the different phases and Rashba couplings such that the bulk gap is comparible. It turns out that this approximatly holds if we pick $m = 7$ and $r_0 = 0$ or $r_0 = 1$ for the QSHI and $m = 9$ and $r_0 = 0$ or $m = 10$ and r_0 for the TCI. The result for the transmission in the bulk gap and a little part of the bulk band is shown in figure 5.4.

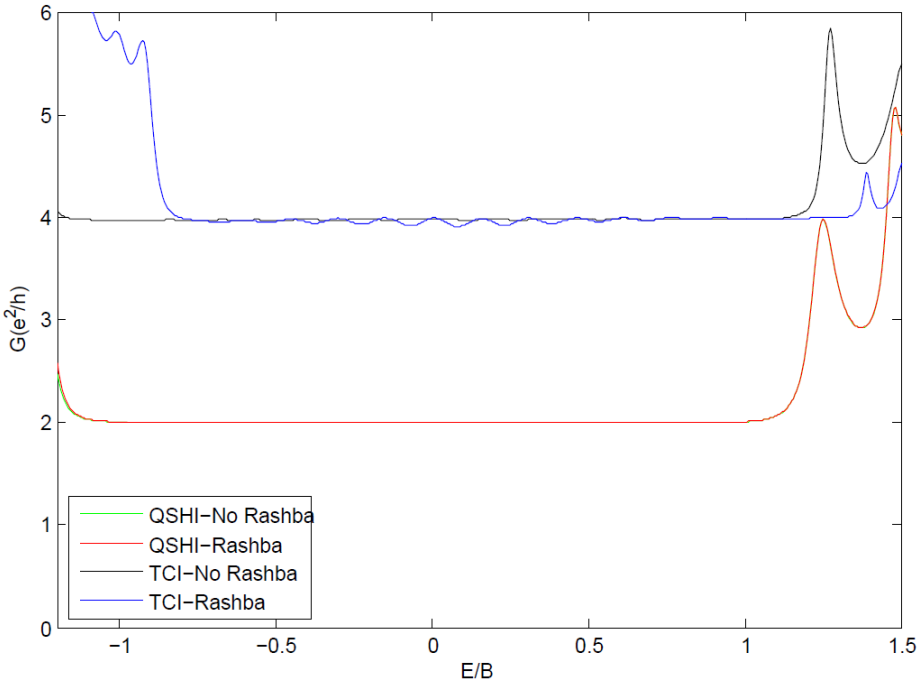

Figure 5.4: The conductance in the bulk gap for the various phases and Rashba couplings. Note that the QSHI-phase with and without Rashba is almost the same here.

In figure 5.4 we can see that the conductance in the TCI bulk gap is indeed $4e^2/h$ and in the QSHI it is $2e^2/h$. Furthermore, the Rashba coupling has essentially no effect on the QSHI, but the TCI conductance is a bit different. We remark that in the TCI-Rashba phase we increased m a bit, in order to keep the total bulk gap width comparible.

If we further compare the results in figure 5.4 with the results in section 3.3 it is notable that in the bulk of the systems in figure 5.4 the transmission is higher (which is not very surprising considering the width of these systems) and there are clear Fabry-Pérot oscillations, while these oscillations

are absent in the bulk gap.[21] This is again a clue that along a given edge there is really "one-way traffic" and there is no interference from reflecting waves. The tiny oscillations in the TCI-Rashba phase indicate again that Rashba coupling may destabilze the TCI boundary conductance.

The best way of finding out whether a system is stable under disorder is of course by performing a simulation. The disorder was implemented by picking at each site[22] a perturbation to the chemical potential such that these perturbations were uniformly distributed between $-w$ and w; we refer to w as the disorder strength. This means that time-reversal symmetry is always maintained and the rotational symmetry of the TCI-phase is maintained on average. The results are shown in figures 5.5 and 5.6.

Figures 5.5 and 5.6 show clearly that the disorder does decrease the conductance from the TCI boundary modes when there is Rashba coupling, while in the absence of coupling the TCI is just as stable as the QSHI (and on the QSHI Rashba coupling has no effect at all). The stability of the TCI without Rashba coupling we find is in agreement with [64] and [65], which were published recently.[23]

We remark that the stability of the system without Rashba coupling in both phases is in agreement with our earlier analysis that the QSHI consists of two decoupled Chern insulators with $C_1^{\uparrow} = -C_1^{\downarrow} = -1$ [61, 66, 9] and the TCI of two decoupled Chern insulators with $C_1^{\uparrow} = -C_1^{\downarrow} = -2$ [9], for which there can be no backscattering in either case. Furthermore, even though Rashba coupling turns the spin polarized modes of the QSHI into helical modes, it will still be protected by time reversal symmetry [62, 63, 67, 68, 9].

We can thus conclude that the TCI phase may be as stable as the QSHI phase, as long as there is no Rashba coupling. In the presence of Rashba coupling the boundary properties of the QSHI remain essentially unchanged, while the boundary properties of the TCI change dramatically, making it unstable against disorder. Even though we only implemented one specific form of Rashba coupling, the arguments for the instability of TCI edge modes should also apply for other forms of spin-orbit coupling, effectively making the realization of a TCI stable against disorder in realistic materials unlikely.

[21] Except in the TCI-Rashba phase, where there remain some tiny oscillations.

[22] We remark that there are four orbitals at each site, so chemical potentials of orbitals that share a site are always perturbed identically.

[23] The work in this section was however caried out independently of these references.

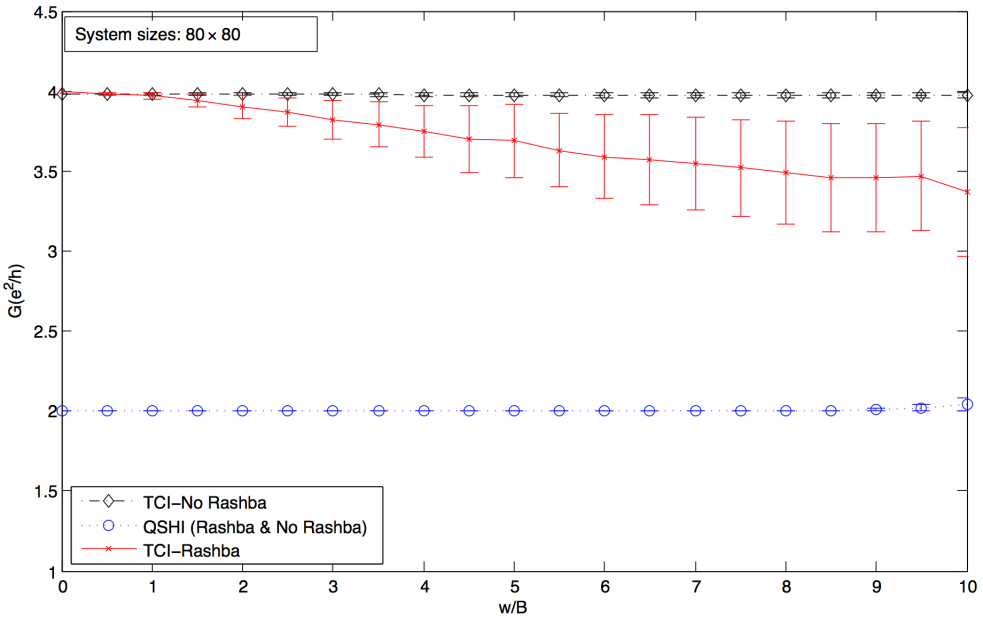

Figure 5.5: The conductance as a function of disorder strength w/B for site disordering. This plot is based on 100 samples of each system.

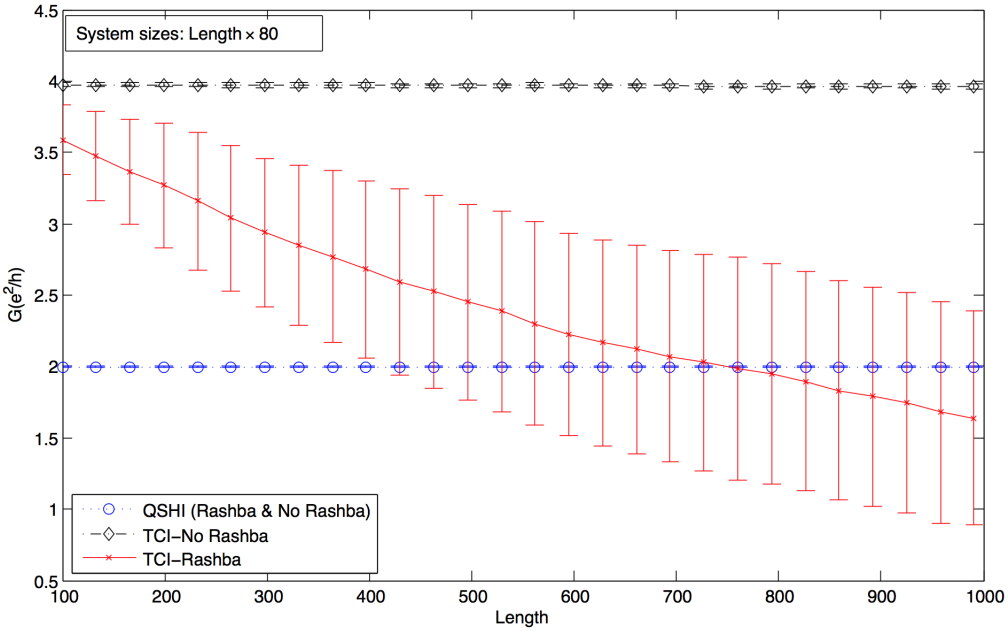

Figure 5.6: The conductance as a function of system length for site disordering with strength $w/B = 5$. This plot is based on 100 samples of each system.

5.2.2.1 Alternative ways of putting disorder into the system

The results in [9] were obtained by disordering sites, which means that chemical potentials of different orbitals on the same site were all shifted by the same amounts. We could also apply disorder differently, for example not disordering each site with the same disorder strength, or by disordering different orbitals at the same site independently. Here we demonstrate what happens if we only disorder the boundary sites of the system (but with disorder strengths much larger than the bulk gap) and also how disordering different orbitals at the same site effects conductance.

If we apply independent disorder to all sites and orbitals, the system without Rashba coupling remains stable and the TCI phase with Rashba coupling remains unstable. There is however a difference in the stability of the QSHI with Rashba coupling: the disorder is now not only a chemical potential disorder, but also a random magnetic field (and a disorder in tuning parameter). The combination of Rashba spin-orbit coupling and time-reversal symmetry breaking by the magnetic field destabilises the QSHI, as can be seen in figure 5.7. This illustrates that for the time-reversal symmetry it is in general necessary to be exactly conserved and not just on average.

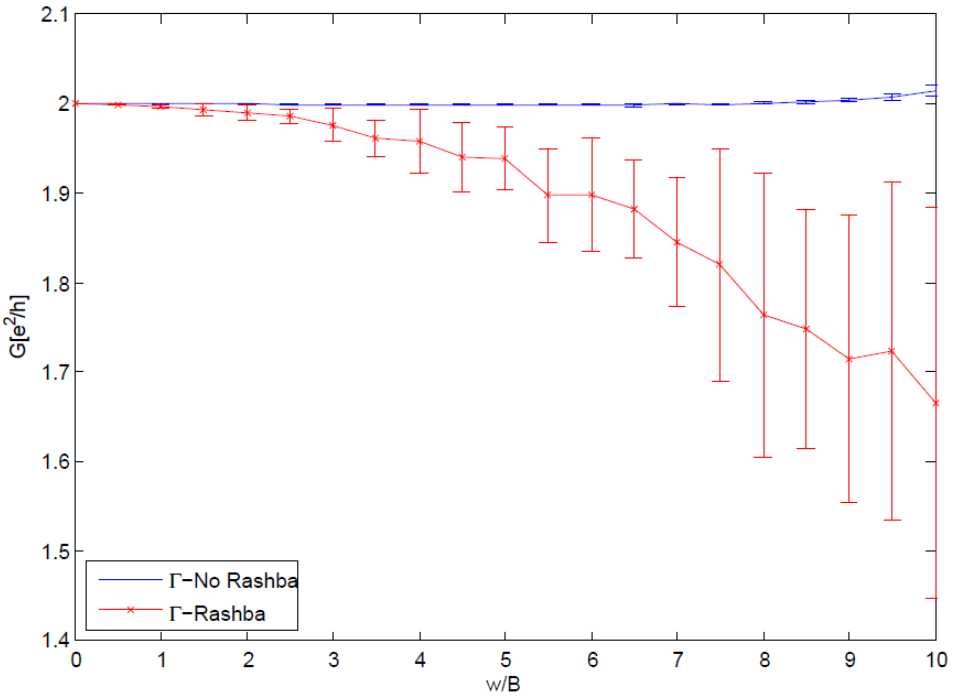

Figure 5.7: The conductance as a function of disorder strength w/B for site and orbital disordering. This plot is based on 100 samples of each system.

The effect of disordering the boundary sites is in fact very simple: with or without Rashba coupling and in both the TCI and QSHI there is no effect on the conductance, even if the disorder is very large. The explanation of this is that it does not matter to the system how wide it is: instead of Dirichlet conditions which forbid hopping out of the system, this will be archived by the very high disorder on the outer sites and we are effectively left with a smaller, but clean system, for which the modes can still propagate along the new boundary.

Chapter 6

Conclusions

We have described how the conductance through electronic systems can be expressed in terms of Green's functions and described and implemented an algorithm to compute the Green's function of an arbitrary lead.

We then discussed different algorithms for computation of this conductance and compared some of them on computational speed and stability. We argued that the best performance might be archived by very sophisticated algorithms and that the recursive Green's function algorithm offered a good compromise between speed, stability and implementability for practical systems. For systems that are very large or for which the conductance needs to be computed extremely often, it may be beneficial to develop and implement faster algorithms than the RGF algorithm.

Finally, we looked at topological insulators, first using an example to review some theory on standard topological insulators. Afterwards we looked at a non-standard topological insulator, where crystalline symmetries also play a role. We used our transport algorithm and an analysis of the system on a cylinder to show that for site disorder the TCI phase is only stable as long as there the spin in the z-direction is conserved. An interesting open problem remains the study of conductance under disorder for other non-standard topological insulators, both in 2D and in 3D.

Appendix A

Some quantum mechanics

Quantum mechanics is the name of a physical theory that is used to describe processes that happen on a small scale. In this appendix we state some parts of the theory that are of importance for our purposes. A much more complete coverage of the theory of quantum mechanics can be found in for example [19].

A.1 The Schrödinger equation

The Schrödinger equation,

$$\left(-\frac{\hbar^2}{2m}\Delta + U(\mathbf{x},t)\right)\Psi(\mathbf{x},t) = i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{x},t), \tag{A.1.0.1}$$

where $\hbar = \frac{h}{2\pi}$ and h is Planck's constant (a constant of nature), m is a mass (a constant that depends on the problem at hand), $U(\mathbf{x},t)$ a given function of position $\mathbf{x}$ and time t that is called *potential* and $\Psi(\mathbf{x},t)$ an unknown function of time and position know as the *state*, is the main equation in quantum mechanics. The operator $\hat{H} := \left(-\frac{\hbar^2}{2m}\Delta + U(\mathbf{x},t)\right)$, the left-hand side of (A.1.0.1) is called the (quantum mechanical) *Hamiltonian*.

If we manage to solve the Schrödinger equation with given initial and boundary conditions, then all properties we may want to measure are encoded in the state $\Psi(\mathbf{x},t)$ (the way to obtain these properties from the state will be discussed in section A.2.2. We will now demonstrate how to compute the state for a simple example and then discuss how Sturm-Liouville theory can be used to handle more general cases.

A.1.1 A particle in a box

Suppose the potential $U(\mathbf{x},t)$ is constant everywhere (without essential loss of generality we take this constant to be 0), we impose zero Dirichlet bound-

ary conditions on an n-dimensional rectangular domain D and we take any initial condition $\Psi(\mathbf{x}, 0) \in L^2(D)$. We can then use separation of variables to reduce this to solving one-dimensional equations and we can rescale and shift the sides of D to reduce (A.1.0.1) to n copies of the following eigenvalue problem

$$-\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \lambda y(x), \tag{A.1.1.1}$$

which is of course solved by $y(x) = \sqrt{\frac{2}{\pi}} \sin(\sqrt{\lambda} x)$.

Since the sine-Fourier basis is a complete orthonormal system on the domain, the solution can be obtained by scaling the solutions back to the original sizes, expanding the initial conditions in terms of the solutions, using that for these eigenfunctions the time-dependence is exponential and multiplying the solutions for the different spatial dimensions to obtain a solution to (A.1.0.1). We remark that the eigenvalues of the eigenfunctions depend quadratically on the wave numbers of the sinusoids.

A.1.2 Sturm-Liouville theory

Suppose the potential is chosen such that (A.1.0.1) is separable. If we separate the variables, we obtain eigenvalue equations of the form

$$\left(-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + U(x)\right)\Psi(x) = E_n \Psi(x), \tag{A.1.2.1}$$

where $U(x)$ is a potential depending on one coordinate and E_n is an (unknown) eigenvalue (with the dimensions of energy). Following [43], we will state some results about eigenstates and -energies.

Suppose that we want to solve A.1.2.1 with Dirichlet boundary conditions ($x \in [a, b] \in \mathbb{R}$ and $\Psi(a) = \Psi(b) = 0$) and assume $U(x)$ is continuous. This is an example of a classical Sturm-Liouville problem. All eigenvalues E_n are real and satisfy $E_n > \inf_{x\in[a,b]} U(x)$. The Schrödinger operator is self-adjoint and there exists an orthonormal system of eigenfunctions.[1]

In practice we often want to solve A.1.2.1 without imposing Dirichlet conditions on a compact interval. The following results are from [43] and [69]. Assume $\int_{-\infty}^{\infty} |U(x)||x|^k \,\mathrm{d}x < \infty, \quad k \in \{0, 1, 2\}$. Then there is a finite (possibly zero) number of discrete eigenvalues $E_n < 0$, for which the eigenstates have finite L^2-norm. These states are known as bound states. There is also a continuous spectrum of eigenvalues with $E > 0$. The eigenstates with these eigenvalues do have finite L^∞-norm, but their L^2-norm is infinite.

[1] See [43] for proofs of these statements.

A.2 Notation and formalism in quantum mechanics

We will now explain how some elements from functional analysis are used in quantum mechanics and introduce a convenient and common notation for the quantum mechanical states, called *Dirac notation*. It will also be explained how quantum mechanical states are related to quantities that could be measured in experiments.

A.2.1 Dirac notation

The quantum mechanical states $\Psi(\mathbf{x}, t)$ we encountered in the previous section are functions of position and time. They might also have other degrees of freedom, like spin, that do not appear explicitly in (A.1.0.1). These quantum mechanical states are elements of some complex L^2-space, which is often referred to by physicists as "the" Hilbert space. This space has of course an inner product $\langle f|g\rangle = \int f^*(\mathbf{x})g(\mathbf{x})\,\mathrm{d}\mathbf{x}$, where the integral is taken over all space and the time is kept fixed.

In the Dirac notation, we drop all dependencies of states on parameters and write a quantum mechanical state $\Psi(\mathbf{x}, t)$ as $|\Psi\rangle$. We introduce the †-operation, called Hermitian conjugate, to take $|\Psi\rangle$ to an element of the dual space $\langle\Psi|$, requiring that $\langle\Psi|\,|\Psi\rangle = \langle\Psi|\Psi\rangle = \int \Psi^*(\mathbf{x})\Psi(\mathbf{x})\,\mathrm{d}\mathbf{x}$. It is clear that † is a bijective operator and that $\langle\Psi|\,|\Psi\rangle = ||\Psi||^2$.

If two states differ only by a multiplication by a complex phase, i.e. if $\langle\Psi'| = e^{i\phi}\langle\Psi|$, for $\phi \in [0, 2\pi[$, we identify them as the same physical state. This means that each physical state is in fact a $U(1)$ principal bundle. Choosing a particular state to represent this equivalence class of states is called *gauge fixing* and transformations that transform between different representations of the same physical state are called *gauge transformations*. We choose to work only with *normalised* states, that is, states for which $\langle\Psi|\,|\Psi\rangle = 1$.

A.2.2 Operators in quantum mechanics

We can also define operators $\hat{O}$ that act on states. We denote this by $\hat{O}\,|\Psi\rangle$. The corresponding operator that acts on $\langle\Psi|$ is denoted by $\hat{O}^\dagger$ and the convention is that it acts from the right, $\langle\Psi|\,\hat{O}^\dagger$. The †-operation transforms the state $\hat{O}\,|\Psi\rangle$ into $\langle\Psi|\,\hat{O}^\dagger$ by definition.

We define the *expectation value* of an operator $\hat{O}$ in a state $|\Psi\rangle$ to be $\langle\Psi|\left(\hat{O}\,|\Psi\rangle\right) = \left(\langle\Psi|\,\hat{O}^\dagger\right)|\Psi\rangle = \langle\Psi|\,\hat{O}\,|\Psi\rangle$, where the equalities are postulated. Let $|\Psi\rangle, |\Psi'\rangle$ be states. If $\langle\Psi'|\,\hat{O}\,|\Psi\rangle = \langle\Psi|\,\hat{O}\,|\Psi'\rangle^*$ for all states, we

call $\hat{O}$ Hermitian. In general we have that $\langle\Psi'|\,\hat{O}\,|\Psi\rangle = \langle\Psi|\,\hat{O}^\dagger\,|\Psi'\rangle^*$. We remark that the expectation value of Hermitian operators is real.

All physical *observables* (i.e. things that can potentially be measured in experiments) are represented by Hermitian operators. Examples of quantum mechanical observables are energy, momentum and position. If we write the state $|\Psi\rangle = \Psi(\mathbf{x}, t)$ as a function of position (and time), the operator representing position $\hat{\mathbf{x}}$ corresponds with multiplication with the position, so $\langle\Psi|\,\hat{\mathbf{x}}\,|\Psi\rangle = \int \Psi^*(\mathbf{x})\mathbf{x}\Psi(\mathbf{x})\,\mathrm{d}\mathbf{x}$. The momentum operator $\hat{\mathbf{p}}$ will be given by $\langle\Psi|\,\hat{\mathbf{p}}\,|\Psi\rangle = \int \Psi^*(\mathbf{x})(i\hbar\nabla)\Psi(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and the expectation value of the energy will (of course) be given by the expectation value of the Hamiltonian, so for the Schrödinger equation (A.1.0.1) this would be $\langle\Psi|\,\hat{H}\,|\Psi\rangle = \int \Psi^*(\mathbf{x})\left(-\frac{\hbar^2}{2m}\Delta + U(\mathbf{x}, t)\right)\Psi(\mathbf{x})\,\mathrm{d}\mathbf{x}$. Although this approach is rather different to the way classical observables would be computed, we remark that, by Ehrenfest's theorem [19], the expectation values of quantum mechanical observables will coincide with the values they would take on in classical mechanics.

We can introduce a special kind of operators known as *projection operators* $\Pi_\Psi := |\Psi\rangle\langle\Psi|$, which project any state in the direction of $|\Psi\rangle$. If $|\Psi\rangle$ is normalised, $\Pi_\Psi\,|\Psi\rangle = |\Psi\rangle$ and if we sum over the projection operators on a complete orthonormal system we obtain the identity operator.

We remark that in many of the systems we consider in this thesis the states are finite-dimensional (column) vectors, so the Hilbert space becomes a finite-dimensional vector space, where the operators are in fact matrices and taking the Hermitian conjugate corresponds to taking the transpose and the complex conjugate of a matrix or vector.

A.2.3 Simultaneous eigenbases

Not all operators that we define for our quantum mechanical system have to commute, but if two operators do commute, they have a *simultaneous eigenbasis*, that is, there exists a basis such that all basis functions are eigenvectors of both the commuting operators. We show this closely following [19].

Suppose the operators $\hat{X}, \hat{Y}$ commute, the eigenfunctions $\{|x_j\rangle\}_{j\in\mathbb{N}}$ (with corresponding eigenvalues x_j) of $\hat{X}$ form a complete orthonormal system and all eigenvalues of $\hat{X}$ are non-degenerate[2].

We can represent $\hat{X}, \hat{Y}$ as infinite matrices with elements $\langle x_j|\,\hat{X}\,|x_k\rangle$, $\langle x_j|\,\hat{Y}\,|x_k\rangle$.

[2] the assumption of non-degeneracy can be lifted, see [19]

It is clear that the matrix representing $\hat{X}$ is diagonal. The following holds

Theorem 3 (Simultaneous eigenbases)**.** $\{|x_j\rangle\}_{j\in\mathbb{N}}$ *are eigenfunctions of* $\hat{Y}$ *with eigenvalues* $\langle x_j|\,\hat{Y}\,|x_j\rangle$*.*

Proof. We first prove the following lemma:

Lemma 1. *The matrix* $\langle x_j|\,\hat{Y}\,|x_k\rangle$ *is diagonal.*

Proof. We use that the operators commute and that the eigenfunctions of $\hat{X}$ form a basis.

$$0 = \langle x_j| \left(\hat{X}\hat{Y} - \hat{Y}\hat{X}\right) |x_k\rangle = (x_j - x_k)\,\langle x_j|\,\hat{Y}\,|x_k\rangle\,,$$

so the matrix representation of $\hat{Y}$ is only non-vanishing for the diagonal elements; using the Kronecker delta we can write $\langle x_j|\,\hat{Y}\,|x_k\rangle = \delta_{j,k}\,\langle x_j|\,\hat{Y}\,|x_j\rangle$. □

We can write $\hat{Y}$ in terms of projection operators and act on an eigenfunction of $\hat{X}$:

$$\hat{Y}\,|x_j\rangle = \sum_{k\in\mathbb{N}} |x_k\rangle\,\langle x_k|\,\hat{Y}\,|x_k\rangle\,\langle x_k|x_j\rangle = \left(\langle x_j|\,\hat{Y}\,|x_j\rangle\right)|x_j\rangle\,, \qquad \text{(A.2.3.1)}$$

which is of course an eigenvalue equation. This proves the theorem, any eigenfunction $|x_j\rangle$ of $\hat{X}$ is an eigenfunction of $\hat{Y}$ with eigenvalue $\langle x_j|\,\hat{Y}\,|x_j\rangle$. □

A.3 The Pauli exclusion principle

Until now in this appendix we have only discussed the quantum mechanical behaviour of single particles. We are however interested in the behaviour of large numbers of (non-interacting) electrons in crystals. An important property of these electrons is that they obey the *Pauli exclusion principle*, which states that two electrons can never be in the same state. We first discuss this exclusion principle and then state some of the consequences for our non-interacting electron systems.

Quantum mechanical particles have the property that they are indistinguishable: all electrons are the same, they just can be in different states. This means that if we have two electrons in different states $|\Psi_1\rangle\,, |\Psi_2\rangle$, the situation after "switching them around" should be physically the same. Hence we should require their combined wave-function to be symmetric or antisymmetric upon changing the labels 1 and 2. Particles which are combined symmetrically are called bosons, particles with antisymmetric combined wave-functions are called fermions. It turns out that for electrons we should

always choose the antisymmetric combination: electrons are fermions.

From the antisymmetry property of the combined wave-functions it immediately follows that the electrons have to be in different states, since their combined wave-function would vanish otherwise. This is then the Pauli exclusion principle. We remark that electrons also have a property known as *spin*, which can take two different values. We only require the entire wave-function, including the spin component, to be antisymmetric, so two electrons with different spin can still have the same spatial wave function.

The main consequence of this is that even non-interacting electrons notice each-others presence, because states that are already filled by other electrons are unavailable. The states themselves do however not change: we can solve the many particle problem by solving the one-particle problem and then filling states until the number of electrons is exhausted.

Appendix B

On disorder in low-dimensional systems

There are important differences in the way disorder affects the conduction properties of one- and two-dimensional systems as opposed to three-dimensional systems. This was discussed in section 2.1.4.2, which is based on [22], [23] and [24]. With our algorithms for the transmission computation, this should be possible to verify.

Our example system is the ideal wire, obeying the discrete Schrödinger equation. We connect it to ideal leads and then disorder the wire, with disorder picked uniformly on the interval $[-w, w]$. We study the transmission as a function of disorder (keeping the length fixed) for a few different energies and then we keep the disorder constant, while increasing the system length. This procedure is quickly implemented using the recursive Green's function algorithm, see section 4.1.

For these 1D systems it can be seen that the conductance decreases as a function of disorder or length and that for sufficiently low disorder this decay becomes asymptotically exponential. This behaviour is indeed the behaviour predicted by the theory, confirming that Anderson localisation does indeed take place in this 1D ordinary metal.

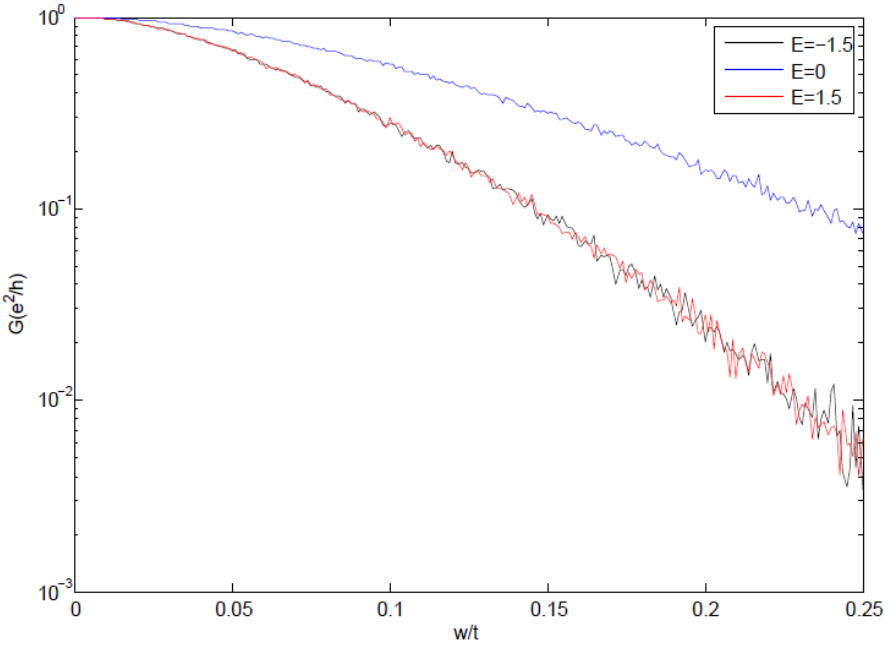

Figure B.1: The conductance as a function of disorder strength w/t for site disordering. This plot is based on 1000 samples of wires with 1000 sites.

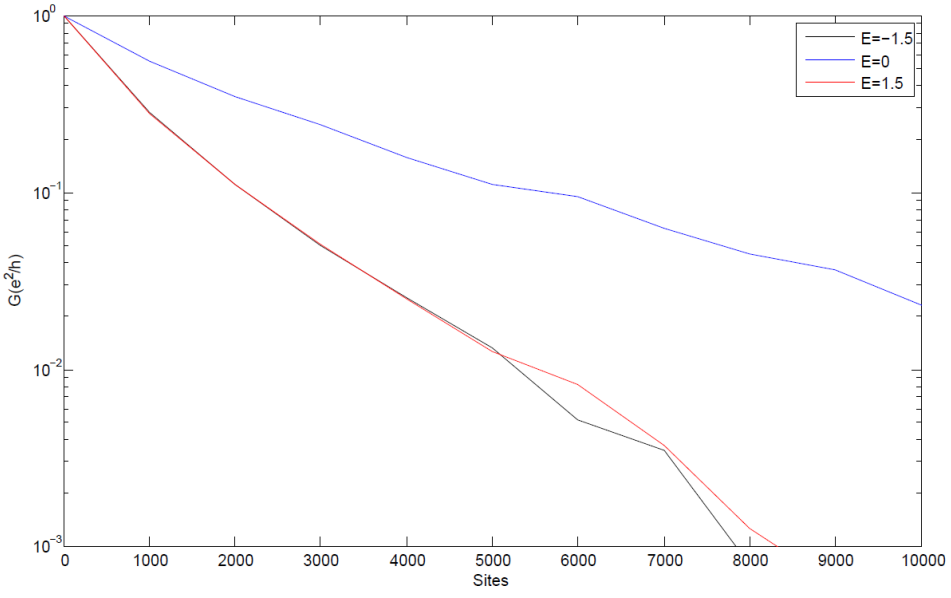

Figure B.2: The conductance as a function of system length for site disordering with strength $w/t = 0.1$. This plot is based on 1000 samples of each system.

Appendix C

Numerical linear algebra

In this thesis many linear algebra computations were performed numerically. In this appendix we describe some features that make numerical linear algebra different from exact linear algebra and we state some algorithms used for our computations, with an analysis of cost and stability of the algorithms. Some references for the material in this appendix are [44], [45] and [53].

C.1 Finite precision linear algebra

One of the main differences between numerical and "ordinary" linear algebra is that in the numeric case all computations done are up to finite precision, so at every step a round-off error creeps in. It will be our goal to make sure the cumulative effect of all these round-off errors remains small in practice. In order to analyse this error accumulation or cancellation, we will introduce a few important theoretical concepts in this section and demonstrate some of them in an analysis of a standard matrix multiplication algorithm.

C.1.1 Finite precision computations

When computing things on a computer, all numbers that are used will be represented by a finite number of bits. This means that numbers can only be represented up to a certain precision and rounding of numbers can be a source of errors.

For the arithmetic, we use the model described in [45], which says that if the numbers x, y are representable in the arithmetic and $\text{op} = +, -, *, /$, then $\frac{\text{computed}(x \text{ op } y)}{\text{exact}(x \text{ op } y)} = (1+\delta)$, with $|\delta| \leq u$ and u a (small) number known as the machine precision. We assume this error estimate to hold for the square root operation too. We remark that for MATLAB the standard data type is the double as defined in the IEEE standard 754 [46]. For this arithmetic, the machine precision $u = 2^{-53} \approx 1.11 * 10^{-16}$ [45], [70].

We also introduce some notation useful in error analysis (γ_n-notation, see [45]): let n be an integer and u the machine precision. Then $\gamma_n := \frac{nu}{1-nu}$. When we encounter γ_n in error bounds, we implicitly assume $nu \ll 1$.

We want to find algorithms such that even with finite precision computations the result we obtain is in some sense close to the answer of the problem we attempt to solve. We call a computation *forward stable* if the computed answer differs only a little from the exact answer. If the computed answer is the exact answer to a slightly perturbed problem, we refer to that as *backward stability*. If we almost solve a slightly perturbed problem, we have *mixed forward-backward stability*. How large the acceptable deviations are of course depends on the problem at hand. [45]

We will be primarily interested in problems with complex numbers. The results in this appendix are derived for real linear algebra, but they all generalize to the complex case, provided we let $\delta \in \mathbb{C}$ and rescale $u \to \gamma_4$. [45]

C.1.2 Matrix norms

We would like to state, using only one number, how "far" two matrices are apart. For this purpose we define some matrix norms. Even though all norms on finite-dimensional spaces are equivalent, in practice some of them are more useful then others.

The matrix norms most commonly used in numerical analysis are the Frobenius norm $||.||_F$

$$||A||_F := (\mathrm{Tr}(A^*A))^{\frac{1}{2}} \tag{C.1.2.1}$$

and matrix norms subordinate to Hölder p-norms

$$||A||_p := \sup_{\mathbf{x}\neq 0} \frac{||A\mathbf{x}||_p}{||\mathbf{x}||_p}. \tag{C.1.2.2}$$

C.1.3 Matrix multiplication

The most common algorithm for matrix multiplication and its analysis are straightforward. The algorithm is given by implementing the standard formula for matrix multiplication. Let A, B be $n \times n$ matrices and define $C = AB$. Then the element $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. To compute all elements of C, we need to compute n^2 of these sums, taking $2n^3$ flops in total.

The error analysis can be performed by considering all the individual com-

putations in the matrix multiplication. This leads [1] to the following stability estimates

$$|C - \hat{C}| \leq \gamma_n |A||B|,$$
$$||C - \hat{C}||_q \leq \gamma_n ||A||_q ||B||_q, \qquad q \in \{1, 2, F\},$$

where $\hat{C}$ is the computed matrix, $|A|$ is the matrix consisting of the absolute values of the elements of A and the norms are the same as those in the previous section.[2]

There are some variations possible on the sketched algorithm for matrix multiplication, mainly in the precise order in which the computations are done and the way (intermediate) results are stored. By choosing the appropriate algorithm, the constants in the stability estimates can be reduced [45] and the speed at which the practical computation is carried out can be increased [44]. However, the theoretical stability and complexity analysis remains the same. Since our purpose here is to enable this theoretical analysis for the algorithms in this thesis and because we can assume the programmers of linear algebra libraries to be aware of these practical issues, there is no need to delve deeper into them here.

We also remark that there exist algorithms for asymptotically faster matrix multiplication (such as Strassen's algorithm [71])[3]. The suspicion is that the best possible complexity of matrix inversion has exponent 2 [73][4]. However, these algorithms are not useful in practice, since their advantage only shows for extremely large matrices and there may be stability problems, so we assume that the cost scales as (constant times) the third power of the matrix dimensionality.

C.1.4 Condition numbers

It is well-known for a system of linear equations $A\mathbf{x} = \mathbf{b}$ there exists a unique solution for all $\mathbf{b}$ if and only if A is non-singular. Numerically, it is easier to solve such an equation if A is "far away" from singularity.

The condition number is a way of expressing this distance from singularity. Suppose A is a non-singular matrix. Then A^{-1} exists and we can define the condition number $\kappa(A) := ||A||||A^{-1}||$, where any norm can be used. The precise value of the condition number of course depends on the specific norm we use and if we want to state something about a condition number

[1] (see [45])
[2] The first of these equations hold element-wise.
[3] the fastest algorithm today scales with an exponent 2.3728639 [72]
[4] this is certainly a minimum, since we should do something with each matrix element

computed with a specific norm $||.||_q$, we write $\kappa_q(...)$. For any matrix A, $\kappa(A) \geq 1$.

Let A be a $n \times n$ complex non-singular matrix. The following theorem expresses the relative distance to singularity as the inverse of the condition number:

Theorem 4 (Gastinel, Kahan)**.** *Let $p \in [1, \infty]$. Then*

$$\kappa_p(A)^{-1} = \min_{\Delta A} \left\{ \frac{||\Delta A||_p}{||A||_p} \middle| A + \Delta A \text{ singular} \right\}. \tag{C.1.4.1}$$

Proof. See [74] or [45]. □

There are many ways [45] to estimate errors for the numerical computation of $A\mathbf{x} = \mathbf{b}$. Here we state one of them, but better estimates can be made. Let $\epsilon > 0$, $A\mathbf{x} = \mathbf{b}$ and $(A + \Delta A)\mathbf{y} = \mathbf{b} + \mathbf{\Delta b}$ (both in exact arithmetic), with $||\Delta A|| \leq \epsilon ||A||$, $||\mathbf{\Delta b}|| \leq \epsilon ||\mathbf{b}||$ and $\epsilon\kappa(A) \leq 1$, then [45]

$$\frac{||x - y||}{||x||} \leq \frac{2\epsilon\kappa(A)}{1 - \epsilon\kappa(A)}. \tag{C.1.4.2}$$

This result says that if A is well-conditioned, then the solution of a linear system is not very sensitive to small perturbations and we can expect algorithms to be normwise mixed forward-backward stable.

Another important result involving condition numbers relates *left-* and *right-residuals* to each other. Let A a non-singular square matrix, X a matrix of the same dimensions[5] and $\hat{\mathbf{1}}$ the appropriate identity matrix. Then the left-residual is defined as $XA - \hat{\mathbf{1}}$ and the right residual as $AX - \hat{\mathbf{1}}$. Since $A\left(XA - \hat{\mathbf{1}}\right)A^{-1} = AX - \hat{\mathbf{1}}$,

$$\kappa(A) \geq \left\{ \frac{||AX - \hat{\mathbf{1}}||}{||XA - \hat{\mathbf{1}}||}, \frac{||XA - \hat{\mathbf{1}}||}{||AX - \hat{\mathbf{1}}||} \right\}, \tag{C.1.4.3}$$

implying that for a well-conditioned matrix A, a good approximate right-inverse will also be a good approximate left-inverse and vice versa.

C.2 Sparse linear systems

In chapter 4 we describe algorithms for the numerical computation of transmission function. The computationally hardest step of these algorithms usually consists of solving many equations of the form $A\mathbf{x} = \mathbf{e_i}$, where A is

[5] This matrix X will in practice be the computed inverse.

a given matrix, $\mathbf{x}$ an unknown vector and $\mathbf{e_i}$ a standard basis vector.[6]

Except for the RGF algorithm, the matrix A is sparse in these computations. The case were A is dense will be treated in the section on LU decomposition. If A is sparse (and in our case with a sparsity pattern consisting of blocks in the upper-left and lower-right corners and some bands), we can either apply a direct method (LU decomposition) or an iterative method (we describe (preconditioned) GMRES, but other algorithms are also possible).

C.2.1 LU decomposition

The idea of the LU decomposition is to split the general system of linear equations $A\mathbf{x} = \mathbf{e_i}$ into a larger but easier system

$$L\mathbf{y} = \mathbf{e_i}$$
$$U\mathbf{x} = \mathbf{y},$$

where L is unit lower triangular, U is upper triangular, $LU = A$ and the vector $\mathbf{y}$ is there for computational purposes only. It can be shown [44] that if A is of size $n \times n$ and for all $k \in \{1, ..., n-1\}$ the $k \times k$ leading principal submatrix[7] is non-singular, there exists a LU decomposition of A. Furthermore [44], if an LU decomposition of a non-singular matrix A exists, it is unique.

We remark that it is very easy to solve a triangular linear system: there will be one linear equation with only one unknown; solving for this unknown and substituting the solution will reduce another equation to a linear equation with only one unknown left etcetera. This forward (for L-matrices) or back (for U-matrices) substitution is numerically very stable and the cost scales as $\mathcal{O}(n^2)$[45].

The factors L and U can be computed using Gauss elimination (with partial pivoting). This standard method is stable in practice[8] and costs $\mathcal{O}(n^3)$ flops. If A is well-conditioned, this method gives small residuals.[44]

The above results hold for full matrices, but if A is sparse, there may be

[6] The inversion step in the RGF algorithm can also be written as solving this equation for all standard basis vectors for a space with the dimension of the slice size. In this way we would compute a numerical right inverse. However, in our implementation the Matlab inv()-routine is used, which computes the numerical left inverse [45]. Since the left- and right-residuals are comparable for well-conditioned matrices and the inv()-routine does use LU decomposition anyway, this subtlety is not of much practical importance.

[7] That is, the matrix formed by taking the elements $a_{ij} \in A$ with $i \leq k$ and $j \leq k$.

[8] The stability estimates contain a "growth factor" for which no satisfactory estimates can be made but which is rather low for many matrices [45]

ways to speed up this procedure. We will look at the particular case where A is banded with band width p. Then the LU decomposition will inherit this band width and can be computed with a cost of $\mathcal{O}(np^2)$[44]. It should be noted that the LU bands will generally be filled from diagonal until the p-th subdiagonal, even if A used to have some empty bands in between. This is known as *fill-in* and it has a negative effect on performance, so it is beneficial to reorder A in a way that minimizes this fill-in. If Gauss elimination with partial pivoting is used, the columns of L will be permuted (but it remains unit lower triangular) and the bandwidth of U doubles [44]. This should however just contribute a constant factor to the computational cost.

C.2.2 Iterative methods

Iterative methods are a way of solving sparse matrix equations approximately. The basic idea is to start with some guess of the solution to the equation, analysing how good a candidate solution it is and then improving this guess, after which the procedure can be repeated with the new guess. For a good method, the number of iterations a candidate solution goes through is small, since then the iterative method may arrive at a reliable solution more efficiently then a direct method. A reason why iterative methods are especially suitable for sparse matrix equations is that sparse matrix-vector multiplications can be carried out very quickly, which means each iteration step can also be done rather fast. A good reference on iterative methods is [50]. In this section we will give a short description on the method we applied in chapter 4, that is, we describe the Generalized Minimal Residuals method (GMRES) and discuss preconditioning to improve convergence.

C.2.2.1 Generalized minimal residuals (GMRES)

GMRES is an example of a *Krylov subspace* method. Krylov subspace methods approximate $\mathbf{x} = A^{-1}\mathbf{b}$ by $\mathbf{x_m} = \mathbf{x_0} + q_{m-1}(A)(\mathbf{b} - A\mathbf{x_0})$, where m is the iteration step, $\mathbf{x_0}$ an initial guess for the solution and q_{m-1} some (cleverly chosen) polynomial of degree $m-1$. We define the initial residual $\mathbf{r_0} := \mathbf{b} - A\mathbf{x_0}$ and the Krylov subspace $\mathcal{K}_m(A, \mathbf{r_0}) := \text{span}(\mathbf{r_0}, A\mathbf{r_0}, ..., A^{m-1}\mathbf{r_0})$. In GMRES, we find the sequence of polynomials $q_0, q_1, ...$ by imposing the condition that the residual at iteration $\mathbf{r_m} := \mathbf{b} - A\mathbf{x_m}$ is orthogonal to the space $A\mathcal{K}_m(A, \mathbf{r_0})$; this is done by a projection step.[50]

For reference, we state a version of the GMRES algorithm from [50] (algorithm 6.9), to which we refer for further details on the algorithm. In the thesis we used the standard Matlab implementation of GMRES, which is slightly different from the algorithm below; the Matlab implementation is more sophisticated but therefore also harder to write down and our rela-

tively simple algorithm is already sufficient for our analysis of the method. We note that in this simple algorithm only the solution after a fixed number of iterations m is given.

GMRES algorithm (Algorithm 6.9 in [50])

- Define initial residual norm $\beta := ||\mathbf{r_0}||_2$ and the initial unit search vector $\mathbf{v_1} := \mathbf{r_0}/\beta$.
- For $j = 1 : m$
  - $\mathbf{w_j} := A\mathbf{v_j}$
  - Orthogonalize the $\mathbf{w_j}$ using Modified Gram-Schmidt (MGS), that is, For $i = 1 : j$
    - $h_{ij} := \mathbf{w_j} \cdot \mathbf{v_i}$
    - $\mathbf{w_j} = \mathbf{w_j} - h_{ij}\mathbf{v_i}$
  - $h_{j+1,j} := ||\mathbf{w_j}||_2$. If $h_{j+1,j} = 0$, the residual is 0 and we can break out of the iteration to compute $\mathbf{x}$.
  - The next unit search vector is defined as $\mathbf{v_{j+1}} := \mathbf{w_j}/h_{j+1,j}$.
- Define the $(m + 1) \times m$ matrix H with entries h_{ij} (and other entries zero) and the matrix V for which the columns are the vectors v_i.
- Find $\mathbf{y_m} := \min_{\mathbf{y} \in \mathbb{R}^m} ||\beta\mathbf{e_1} - H\mathbf{y}||_2$ and compute $\mathbf{x_m} := \mathbf{x_0} + V\mathbf{y_m}$.

The computational cost of the GMRES algorithm lies mainly in the (sparse) matrix-vector multiplications (for square d-dimensional systems of length L such as described in chapter 4 (and with A the inverse Green's function), this costs $\mathcal{O}(mL^d) + \mathcal{O}(mL^{2(d-1)})$ flops) and the orthogonalization (which costs $\mathcal{O}(m^2L^d)$ flops in our systems). Other steps, such as the computation of the minimizer, can be implemented relatively cheaply [50]. We remark that for our practical problem we would have to apply the GMRES-procedure L^{d-1} times, for different right hand sides.

C.2.2.2 Preconditioning

From the above cost estimation it is clear that in order for the GMRES-algorithm to perform well, the residual should become sufficiently small after a low number of iterations. The amount of iterations needed will depend strongly on properties of the matrix A. The idea of preconditioning is finding non-singular matrices M_1, M_2, such that $M_1^{-1}AM_2^{-1}M_2\mathbf{x} = M_1^{-1}\mathbf{b}$ is a more suitable equation for tackling with GMRES and such that equations of the form $M_1\mathbf{y_1} = \mathbf{c_1}$ and $\mathbf{y_2} = AM_2^{-1}\mathbf{c_2}$ (for given $\mathbf{c_1}, \mathbf{c_2}$) can be solved

easily.[9]

It is in general not easy to find good preconditioners, since the requirement that the equation is turned into a "good" equation for GMRES means we want the matrices M_1, M_2 to be chosen such that the spectrum of $M_1^{-1}AM_2^{-1}$ is clustered around some non-zero value and its eigenvectors are mutually almost orthogonal. However, to make equations involving M_1 and M_2 easy to solve, these matrices should be as simple as possible. A third requirement is that M_1 and M_2 can be obtained using a procedure that is relatively simple.

The problem of finding the right preconditioner often differs per problem. It has been adressed in the literature for a large number of problems [50, 55, 56, 54, 53] and there are some standard techniques that may work for a particular problem. Among these standard techniques are *incomplete LU factorizations* (ILU), in which we search for a lower triangular matrix L and an upper triangular matrix U such that LU approximates A and the matrices L, U satisfy certain constraints. The simplest variant of ILU is called ILU(0); we use it in chapter 4. It consists of doing an LU decomposition with the constraint on the matrices L, U that they have a zero entry whenever A has a zero entry (so we ignore any fill-in). Although this is in general a crude approximation, the matrices L, U and are easy to compute and equations involving them are quickly solvable.

[9] A system in which M_1 is the identity matrix is called *right-preconditioned*, a system in which M_2 is the identity matrix is called *left-preconditioned* and a system in which neither of them is the identity matrix is called *split-preconditioned*.

Bibliography

[1] A. Kitaev. Periodic table for topological insulators and superconductors. *AIP Conference Proceedings*, 1134(1):22–30, 2009.

[2] S. Ryu; A.P. Schnyder; A. Furusaki and A.W.W. Ludwig. Topological insulators and superconductors: ten-fold way and dimensional hierarchy. *New J. Phys.*, 12:065010, 2010.

[3] A.P. Schnyder; S. Ryu; A. Furusaki and A.W.W. Ludwig,. Classification of topological insulators and superconductors in three spatial dimensions. *Phys. Rev. B*, 78:195125, 2008.

[4] L. Fu. Topological Crystalline Insulators. *Phys. Rev. Lett.*, 106:106802, 2011.

[5] R.J. Slager; A. Mesaros; V. Juričić and J. Zaanen. The space group classification of topological band insulators. *Nature Physics*, 9:98–102, 2013.

[6] C.K. Chiu; H. Yao and S. Ryu. Classification of topological insulators and superconductors in the presence of reflection symmetry. *Phys. Rev. B*, 88:075142, 2013.

[7] T. Morimoto and A. Furusaki. Topological classification with additional symmetries from Clifford algebras. *Phys. Rev. B*, 88:125129, 2013.

[8] K. Shiozaki and M. Sato. Topology of crystalline insulators and superconductors. *Phys. Rev. B*, 90:165114, 2014.

[9] B.M. de Leeuw; C.S. Küppersbusch; V. Juričić and L. Fritz. tudy of stability of topological crystalline insulators against disorder. *ArXiv e-prints*, 2014. http://arxiv.org/abs/1411.0255.

[10] L.V. Keldysh. Diagram technique for nonequilibrium processes. *Sov. Phys. JETP*, 20:1018–1026, 1965.

[11] L. P. Kadanoff and G. Baym. *Quantum Statistical Mechanics*. Benjamin, New York, 1962.

[12] D.S. Fisher and P.A. Lee. Relation between conductivity and transmission matrix. *Phys. Rev. B*, 23:6851–6854, 1981.

[13] S. Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, Cambridge, 1997.

[14] M. Wimmer. *Quantum transport in nanostructures: From computational concepts to spintronics in graphene and magnetic tunnel junctions*. PhD thesis, Universität Regensburg, 2008.

[15] D.L. Sidebottom. *Fundamentals of Condensed Matter and Crystalline Physics*. Cambridge University Press, Cambridge, 2012.

[16] G. Voronoi. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *Journal für die Reine und Angewandte Mathematik*, 133:97–178, 1908.

[17] F. Bloch. Über die Quantenmechanik der Elektronen in Kristallgittern. *Z. Physik*, 52:555–600, 1928.

[18] http://en.wikipedia.org/wiki/bloch_wave. Accessed 15 april 2014.

[19] J.J. Sakurai and J.J. Napolitano. *Modern Quantum Mechanics*. Pearson Education, 2007.

[20] A. Altland and L. Fritz. Primer on topological insulators. Lecture notes.

[21] U.M. Ascher. *Numerical Methods for Evolutionary Differential Equations*. SIAM, Philadelphia, 2008.

[22] P.W. Anderson. Absence of Diffusion in Certain Random Lattices. *Physical Review*, 109:1492–1505, 1958.

[23] D.J. Thouless. Electrons in disordered systems and the theory of localization. *Physics Reports*, 13:93 – 142, 1974.

[24] E. Abrahams; P. W. Anderson; D.C. Licciardello and T.V. Ramakrishnan. Scaling theory of localization: Absence of quantum diffusion in two dimensions. *Phys. Rev. Lett.*, 42:673–676, 1979.

[25] R. Landauer. Spatial Variation of Currents and Fields Due to Localized Scatterers in Metallic Conduction. *IBM Journal of Research and Development*, 1:223–231, 1957.

[26] M. Büttiker. Absence of backscattering in the quantum Hall effect in multiprobe conductors. *Phys. Rev. B*, 38:9375, 1988.

[27] H.D. Cornean; A. Jensen and V. Moldoveanu. A rigorous proof of the Landauer-Büttiker formula. *Journal of mathematical physics*, 46:042106, 2005.

[28] G.D. Mahan. *Many-Particle Physics*. Kluwer Academic/Plenum Publishers, New York, 2000.

[29] H. Brezis. *Functional Analysis, Sobolev Spaces And Partial Differential Equations*. Springer-Verlag New York Inc., New York, 2010.

[30] L.C. Evans. *Partial Differential Equations*. American Mathematical Society, 1997.

[31] J.J. Duistermaat and J.A.C. Kolk. *Distributions: Theory and Applications*. Birkhäuser, 2010.

[32] R.L. Schilling. *Measures, Integrals And Martingales*. Cambridge University Press, Cambridge, 2005.

[33] S. Lang. *Complex Analysis*. Springer-Verlag New York Inc., New York, 1999.

[34] D.A. Ryndyk. Coherent transport: Green function method. The 2d seminar of the course "Primer of quantum transport at nanoscale" WS 2012-2013, 2012.

[35] W.P. Su; J.R. Schrieffer and A.J. Heeger. Solitons in Polyacetylene. *Physical Review Letters*, 25:1698–1701, 1979.

[36] Y. Zhang; M. Han and L. Shen. Conductance oscillation in graphene-nanoribbon-based electronic Fabry–Perot resonators. *Physica B: Condensed Matter*, 405:1168 – 1171, 2010.

[37] C.W. Groth; M. Wimmer; A.R. Akhmerov and X. Waintal. Kwant: a software package for quantum transport. *ArXiv e-prints*, 2013. http://arxiv.org/abs/1309.2926.

[38] C.W. Groth; M. Wimmer; A.R. Akhmerov and X. Waintal. Kwant 1.0.0 documentation: 2.2. First steps: setting up a simple system and computing conductance. Accessed 5 may 2014.

[39] P.S. Krstic; X.G. Zhang and W.H. Butler. Generalized conductance formula for the multiband tight-binding model. *Phys. Rev. B*, 66:205319, 2002.

[40] A.R. Rocha; V.M. Garcia-Suarez; S. Bailey; C. Lambert; J. Ferrer and S. Sanvito. Spin and molecular electronics in atomically generated orbital landscapes. *Phys. Rev. B*, 73:085414, 2006.

[41] 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.

[42] A. Umerski. Closed-form solutions to surface Green's functions. *Phys. Rev. B*, 55:5266–5275, 1997.

[43] J.J. Duistermaat and W. Eckhaus. *Analyse van Gewone Differentiaalvergelijkingen*. Epsilon Uitgaven, Utrecht, 2009.

[44] G.H. Golub and C.F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, 1996.

[45] N.J. Highham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, 1996.

[46] The MathWorks, Inc. *MATLAB R2012B Documentation*, 1994-2012.

[47] A. MacKinnon. The calculation of transport properties and density of states of disordered solids. *Z. Phys. B*, 59:385–390, 1985.

[48] P. A. Lee and D. S. Fisher. Anderson localization in two dimensions. *Phys. Rev. Lett.*, 47:882–885, 1981.

[49] D. J. Thouless and S. Kirkpatrick. Conductivity of the disordered linear chain. *J. Phys. C: Solid State Phys*, 14:235–245, 1981.

[50] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, 2nd edition, 2003.

[51] P.S. Drouvelis; P. Schmelcher and P. Bastian. Parallel implementation of the recursive Green's function method. *Journal of Computational Physics*, 215:741–756, 2006.

[52] P. Amestoy; M. Bremond; A. Buttari; A. Guermouche; G. Joslin; J.Y. L'Excellent; F.H. Rouet; B. Ucar and C. Weisbecker. Mumps users' guide. http://mumps.enseeiht.fr/doc/userguide_4.10.0.pdf. Accesed 4 december 2014.

[53] A.H. Sheikh; D. Lahaye and C. Vuik. On the convergence of shifted Laplace preconditioner combined with multilevel deflation. *Numer. Linear Algebra Appl.*, 20:645–662, 2013.

[54] Y.A. Erlangga. Advances in iterative methods and preconditioners for the helmholtz equation. *Archives of Computational Methods in Engineering*, 15:37–66, 2008.

[55] Y.A. Erlangga and R.Nabben. On a multilevel Krylov method for the Helmholtz equation preconditioned by shifted Laplacian. *Electronic Transactions on Numerical Analysis (ETNA)*, 31:403–424, 2008.

[56] Y.A. Erlangga; C. Vuik and C.W. Oosterlee. On a class of preconditioners for solving the Helmholtz equation. *Appl. Numer. Math.*, 50:409–425, 2004.

[57] J.E. Moore; Y. Ran and X.G. Wen. Topological surface states in three-dimensional magnetic insulators. *ArXiv e-prints*, 2008. http://arxiv.org/abs/0804.2527v2.

[58] F.W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer, New York, 1983.

[59] J.E. Avron; R. Seiler and B. Simon. Homotopy and Quantization in Condensed Matter Physics. *Phys. Rev. Lett.*, 51:51–53, Jul 1983.

[60] L.E.J. Brouwer. Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen*, 71:97–115, 1911.

[61] 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.

[62] E. Wigner. Über die Operation der Zeitumkehr in der Quantenmechanik. *Nachr. Akad. Ges. Wiss. Göttingen*, 31:546–559, 1932.

[63] B.A. Bernevig; T.L. Hughes and S.C. Zhang. Quantum Spin Hall Effect and Topological Phase Transition in HgTe Quantum Wells. *Science*, 314(5806):1757–1761, 2006.

[64] H. Jiang; H. Liu; J. Feng; Q. Sun and X.C. Xie. Transport Discovery of Emerging Robust Helical Surface States in $Z_2 = 0$ Systems. *Phys. Rev. Lett.*, 112:176601, 2014.

[65] M. Ezawa. Electrically tunable conductance and edge modes in topological crystalline insulator thin films: minimal tight-binding model analysis. *New J. Phys.*, 16:065015, 2014.

[66] D.J. Thouless; M. Kohmoto; M.P. Nightingale and M. den Nijs. Quantized Hall Conductance in a Two-Dimensional Periodic Potential. *Phys. Rev. Lett.*, 49:405–408, 1982.

[67] C.L. Kane and E.J. Mele. Z_2 Topological Order and the Quantum Spin Hall Effect. *Phys. Rev. Lett.*, 95:146802, 2005.

[68] C.L. Kane and E.J. Mele. Quantum Spin Hall Effect in Graphene. *Phys. Rev. Lett.*, 95:226801, 2005.

[69] W. Eckhaus and A. van Harten. *The Inverse Scattering Transformation and the Theory of Solitons.* North Holland, Amsterdam, 1981.

[70] IEEE Standard for Floating-Point Arithmetic. *IEEE Std 754-2008*, pages 1–70, 2008.

[71] V. Strassen. Gaussian elimination is not optimal. *Numerische Mathematik*, 13:354–356, 1969.

[72] F. Le Gall. Powers of Tensors and Fast Matrix Multiplication. *ArXiv e-prints*, 2014. http://arxiv.org/abs/1401.7714.

[73] H. Cohn; R. Kleinberg; B. Szegedy and C. Umans. Group-theoretic algorithms for matrix multiplication. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005).*, pages 379–388, 2005.

[74] W.M. Kahan. Numerical linear algebra. *Canadian Mathematical Bulletin*, 9:757–801, 1966.