UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1441

Topological band theory and Majorana fermions

With focus on self-consistent lattice models

KRISTOFER BJÖRNSON

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016

ISSN 1651-6214
ISBN 978-91-554-9728-6
urn:nbn:se:uu:diva-305212

Dissertation presented at Uppsala University to be publicly examined in Häggsalen, Ångströmslaboratoriet, Lägerhyddsvägen 1, Uppsala, Friday, 9 December 2016 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: professor Andrei Bernevig.

Abstract
Björnson, K. 2016. Topological band theory and Majorana fermions. With focus on self-consistent lattice models. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1441. 144 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9728-6.

One of the most central concepts in condensed matter physics is the electronic band structure. Although band theory was established more than 80 years ago, recent developments have led to new insights that are formulated in the framework of topological band theory. In this thesis a subset of topological band theory is presented, with particular focus on topological superconductors and accompanying Majorana fermions. While simple models are used to introduce basic concepts, a physically more realistic model is also studied intensely in the papers. Through self-consistent tight-binding calculations it is confirmed that Majorana fermions appear in vortex cores and at wire end points when the superconductor is in the topologically non-trivial phase. Many other properties such as the topological invariant, experimental signatures in the local density of states and spectral function, unconventional and odd-frequency pairing, the precense of spin-polarized currents and spin-polarization of the Majorana fermions, and a local π-phase shift in the order parameter at magnetic impurities are also investigated.

Keywords: Topology, Majorana, superconductivity, material physics, numerical calculations, tight-binding, mean-field

Kristofer Björnson, Department of Physics and Astronomy, Materials Theory, Box 516, Uppsala University, SE-751 20 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-554-9728-6
urn:nbn:se:uu:diva-305212 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-305212)

Dedicated to you

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Vortex states and Majorana fermions in spin-orbit coupled semiconductor-superconductor hybrid structures**
K. Björnson and A. M. Black-Schaffer
*Phys. Rev. B **88**, 024501 (2013)*

II **Skyrmion spin texture in ferromagnetic semiconductor-superconductor heterostructures**
K. Björnson and A. M. Black-Schaffer
*Phys. Rev. B **89**, 134518 (2014)*

III **Probing vortex Majorana fermions and topology in semiconductor/superconductor heterostructures**
K. Björnson and A. M. Black-Schaffer
*Phys. Rev. B **91**, 214514 (2015)*

IV **Currents Induced by Magnetic Impurities in Superconductors with Spin-Orbit Coupling**
S. S. Pershoguba, K. Björnson, A. M. Black-Schaffer, and A. V. Balatsky
*Phys. Rev. Lett. **115**, 116602 (2015)*

V **Spin-polarized edge currents and Majorana fermions in one- and two-dimensional topological superconductors**
K. Björnson, S. S. Pershoguba, A. V. Balatsky, and A. M. Black-Schaffer
*Phys. Rev. B **92**, 214501 (2015)*

VI **Majorana fermions at odd junctions in a wire network of ferromagnetic impurities**
K. Björnson and A. M. Black-Schaffer
*Phys. Rev. B **94**, 100501(R) (2016)*

VII **Superconducting order parameter π-phase shift in magnetic impurity wires**
K. Björnson, A. V. Balatsky, and A. M. Black-Schaffer

Submitted to Phys. Rev. B

Additional paper, not included in the thesis:

VIII **Solid state Stern-Gerlach spin-splitter for magnetic field sensoring, spintronics, and quantum computing.**
K. Björnson and A. M. Black-Schaffer
Manuscript, arXiv:1509.05266

Reprints were made with permission from the publishers.

My contributions

In all papers except Paper IV and V, I had the main responsibility for carrying out the calculations, analyzing the results, and writing the manuscript. For Paper IV, I carried out the self-consistent calculations presented in Fig. 3 and the supplemental material, while in Paper V I had responsibility for everything except the Ginzburg-Landau calculations in Appendix A.

Contents

[1] Part of the discussion in Section 4.2 and 4.3 runs in parallel with the discussion in the section "Particle-hole symmetry" in Paper III.

[2] This section is particularly hard to read due to the explicit use of multiple notations for both bases and vector component labels. The reader is therefore advised to read the paragraph with pen and paper, and consult Fig. 4.1.

[3] Part of the discussion in this section runs in parallel with the discussion in the section "Majorana spin-polarization and block edge currents" in Paper V.

Part I:
Introduction

1. Introduction

The application of topology to physics has a long history, with early roots in work connected to hydrodynamics and aether theory, through the work of Helmholtz, Thomson, Maxwell, and Tait [1]. Topology has since then found applications in fields such as the study of defects in both condensed matter systems [2, 3] and cosmological models [4], the study of Fermi surfaces in superfluids [5], and in various applications of topological field theories [6, 7].

With the discovery and subsequent theoretical explanation of the quantum Hall effect, yet another application of topology entered the stage [8–10]. Quantum Hall systems are especially interesting, because they host quantized edge currents, so precisely quantized that they among other things have been proposed as a foundation for a resistance standard [11]. As a tool for calculations, the first Chern (or TKNN) number, which until then had been important in particle physics [6], became useful also for condensed matter systems. It became clear that the Chern number, which can be calculated from the samples bulk properties, is directly related to the number of edge channels [10].

Work generalizing the quantum Hall effect eventually culminated with the prediction and subsequent experimental discovery of topological insulators [12–19]. Topological insulators are materials that are insulating in the bulk, but like the quantum Hall state have robust edge states [20–22]. This eventually sparked a wave of proposals for the closely related concept of topological superconductors, which in recent years have attracted a considerable interest because it is predicted that they host Majorana fermions [23–31].

Majorana fermions are particles that arise as solutions to an alternative relativistic equation to the ordinary Dirac equation [32]. No fundamental particles have so far been confirmed to be Majorana fermions, but it is predicted that quasi-particles formally analogous to their high energy siblings can arise in superconductors [33]. The Majorana fermions are of interest in themselves for purely fundamental reasons, but also from a practical point of view. Namely, Majorana fermions have been proposed as building blocks of so called topological quantum computers [34].

While early investigations into topological superconductors was based on highly exotic material properties, recent research have indicated that they can be engineered from building blocks that are comparatively ubiquitous in nature. In partiular, they can be constructed out of materials or combination of materials exhibiting s-wave superconductivity, magnetism, and Rashba spin-orbit interaction. It has been predicted that this kind of setup can host Majorana fermions in superconductor vortex cores and on wire end points [25–30, 35–37]. In this Thesis, which is an extension of an earlier Licentiate Thesis, the

necessary background in topology, topological band theory, and superconductivity that is needed to understand these kind of systems is presented in Chapter 2-4. An introduction to topological superconductivity and the system itself is given in Chapter 5. Further, in Chapter 6-8 some related method development is outlined, while research results presented in Paper I-VII are summarized in Chapter 9.

Part II:
Theoretical background

2. Differential geometry, topology, and fiber bundles

Traditional phases of matter are classified based on their symmetries and order parameters, using Landau's theory of phase transitions [38]. In contrast, topological insulators and superconductors are classified based on topology: a material can be in a topologically trivial or non-trivial phase [20, 21], but what does this mean? For the system we are interested in, the short answer is that it is described by a Hamiltonian that is associated with a non-trivial Chern number. However, the Chern number is an abstract concept, and the statement makes little sense without a proper background in the theory of fiber bundles. In this chapter, we therefore give a short introduction to differential geometry, topology, and fiber bundles, which leads to the definition of the Chern number in the form that is useful to us. A thorough treatment is beyond the scope of this thesis and the aim is therefore not to give a rigorous treatment here. Rather, the focus is at covering enough material to provide a conceptual understanding of the main ideas, and for additional information we refer to the references [6, 39–45].

2.1 Differential geometry

2.1.1 Manifold and tangent space

Differential geometry is a topic in mathematics concerned with the description of manifolds such as for example lines, surfaces, and volumes using calculus. In particular, Riemannian geometry generalizes the concepts of Euclidean geometry, allowing for a systematic study of geometries other than the traditionally flat one. As an example, we consider the sphere. Embedded in three-dimensional Euclidean space, it can be parametrized as

$$\mathbf{r} = (x, y, z) = (r\sin(\theta)\cos(\phi), r\sin(\theta)\sin(\phi), r\cos(\theta)) \,. \tag{2.1}$$

To each point of the sphere, it is further possible to associate a tangent plane, which is spanned by the vectors

$$\frac{\partial \mathbf{r}}{\partial \theta} = (r\cos(\theta)\cos(\phi), r\cos(\theta)\sin(\phi), -r\sin(\theta)) \,, \tag{2.2}$$

$$\frac{\partial \mathbf{r}}{\partial \phi} = (-r\sin(\theta)\sin(\phi), r\sin(\theta)\cos(\phi), 0) \,. \tag{2.3}$$

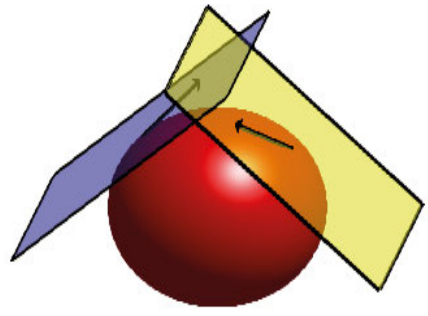

Figure 2.1. In differential geometry, it is important to consider both the manifold and its tangent planes. The manifold is the space parametrized by coordinates, in this case the sphere which is parametrized by the coordinates (θ, ϕ). The tangent spaces are on the other hand Euclidean spaces that are tangent to the manifold. There is one tangent space associated with each point of the manifold, and each tangent space is the space in which tangent vectors based at that point take values.

Two such planes are visualized in Fig. 2.1. Note that this means that we are not only considering the sphere, but also an infinite number of tangent planes: one tangent plane for each point on the sphere. This construction is not explicitly needed in Euclidean geometry, as all tangent planes are parallel[1]. The base space and tangent space can therefore be thought of as being the same. However, in differential geometry it is important to differentiate between the base space manifold and the tangent spaces. Points in the manifold live in the base space, while vectors live in that particular tangent space that is attached to the manifold at the point where the vector has its base. The manifold can in general have any dimension and shape, but the tangent spaces are always Euclidean spaces with the same number of dimensions as the manifold.

2.1.2 Metric and connection

In Euclidean geometry, it is possible to move from one point in the manifold to another along a vector. Since the vectors no longer lie within the manifold itself in Riemannian geometry, it may seem like this possibility is lost once more general spaces are considered. However, we note that vectors can be considered to lie within the manifold as long as they are infinitesimal. The reason is that the component perpendicular to the manifold goes to zero as the square of the infinitesimal length, while the parallel component only goes to zero linearly. It is therefore still possible to move from one point in the manifold to another following tangent vectors, as long as we take a series of

[1] Think for example of a two-dimensional plane, all tangent vectors lie inside that same plane.

infinitesimal steps along the surface. Each step moves us to a nearby point, at which a new tangent vector in the tangent space attached to that point leads us on to the next point. It is in fact even possible to follow a vector that is not infinitesimal. To do so we just need to take infinitesimal steps, carrying the arrow with us, at each step recalibrating the arrow, so that it moves into the tangent space at the new point at which we currently stand.

For a clarifying analogy, imagine a sign here on Earth pointing to a city 500 km away. If taken literally, this arrow points into the stratosphere, as the earth curves away under our feet. However, we may carry the arrow with us, walking in its direction and making sure that we do not twist the arrow in the plane of the ground, but tiliting it to make it always stay horisontal. 500km later we would reach the city. This movement and recalibration of the arrow corresponds to moving the tangent vector from one tangent space into another. It is an example of what is called parallel transport, a concept that we will come back to shortly.

We have seen that we can move from one point to another in the manifold by following infinitesimal tangent vectors. It is therefore clear that the length of these vectors are of interest to us, if we want to measure distances in the manifold. Therefore consider an infinitesimal vector at some coordinate (θ, ϕ), which is associated with a change $(d\theta, d\phi)$ in the coordinates on the sphere[2]

$$dx = r\left(\cos(\theta)\cos(\phi)d\theta - \sin(\theta)\sin(\phi)d\phi\right), \tag{2.4}$$

$$dy = r\left(\cos(\theta)\sin(\phi)d\theta + \sin(\theta)\cos(\phi)d\phi\right), \tag{2.5}$$

$$dz = -r\sin(\theta)d\theta. \tag{2.6}$$

The length of this vector is given by

$$\sqrt{dx^2 + dy^2 + dz^2} = \sqrt{g_{\theta\theta}d\theta^2 + \left(g_{\theta\phi} + g_{\phi\theta}\right)d\theta d\phi + g_{\phi\phi}d\phi^2}, \tag{2.7}$$

where $g_{\theta\theta} = r^2, g_{\theta\phi} = g_{\phi\theta} = 0$ and $g_{\phi\phi} = r^2\sin^2(\theta)$, and $g_{\mu\nu}$ is known as a metric tensor. The purpose of the metric tensor is to enable us to write down quadratic forms in the manifold coordinates, which allows us to calculate the square of (infinitesimal) distances on the manifold. This is a generalization of how the Pythagorean theorem allows us to write down a quadratic form for the square of the Cartesian coordinates in Euclidean geometry. In fact, when using Cartesian coordinates to describe the Euclidean plane, the metric tensor is given by $g^{(E)}_{xx} = g^{(E)}_{yy} = 1, g^{(E)}_{xy} = g^{(E)}_{yx} = 0$.

The metric tensor is in itself a very important construct in differential geometry, allowing us to measure distances, take scalar products, and many other things. However, for our purposes there is one property that is of foremost interest. To explain this we return to the discussion above about parallel transport, where we imagined transporting a vector along a path. As the vector is

[2] These are obtained from the total differentials of the form $dx = \frac{\partial x}{\partial\theta}d\theta + \frac{\partial x}{\partial\phi}d\phi$.

moved, it is continuously re-calibrated, such that it loses the component that is perpendicular to the surface. Through this process, the vector remains inside the respective tangent space at each new point along the path. It turns out that the metric contains enough information to uniquely relate any vector in the tangent space at one point, to the correct tangent vector at a point infinitesimally close. That is, there exists a construct called a connection, which can be computed from the metric, and which facilitates parallel transport. When given any tangent vector and a direction in which to move, the connection tells us which tangent vector corresponds to it in the tangent space at the nearby point. Note that such a construct takes two vectors and gives back one. The first vector is the vector that is to be parallel transported, while the second is a vector that tells us in which direction to transport it. The vector it gives back is that vector in the nearby tangent space, which according to the choice of connection is parallel to the original vector. When the manifold is embedded in a higher dimensional manifold, such as for the sphere embedded in the three-dimensional Euclidean space, the tangent vector obtained through the connection coincides with the tangent vector obtained by "move-and-tilt". However, the connection eliminates the requirement of an embedding manifold, and allows for the concept of parallel transport to be generalized further.

Technically, the connection is in fact only almost what was described above. It turns out that it is useful to divide the object described above into two components, the identity and the connection. The identity will simply transform a vector with components (x, y) at (θ, ϕ) to the vector with the same components at the nearby point. What the connection does is to provide a correction that takes into account that the underlying coordinate system is not necessarily built up of parallel lines. In general, when the connection is constructed from a particular metric, it is known as a Levi-Civita connection, and is given by[3] [6, 46]

$$\Gamma^{\lambda}{}_{\mu\nu} = \frac{1}{2} g^{\lambda\rho} \left(\partial_\nu g_{\rho\mu} + \partial_\mu g_{\rho\nu} - \partial_\rho g_{\mu\nu} \right). \tag{2.8}$$

Here the Einstein convention is assumed, where summation over indices appearing both in the super- and subscripts is implied. Just like the metric can be taken as the starting point for a geometry, without the requirement of an embedding space, it is also possible to take the connection itself as a starting point for a geometry. A geometry therefore does not necessarily have to have a metric as long as a connection is given, something that will be used to generalize the geometrical concept from Riemannian geometry to fiber bundles later in this chapter.

A simple way to understand why parallel transport may imply the change in vector components follows from imagining the ordinary Euclidean plane with polar coordinates (r, θ). Consider the two long blue arrows based at $(3, 0)$ and

[3] Strictly speaking, the Levi-Cevita connection in fact only follows as the unique choice of connection once metricity and zero torsion is assumed [6].

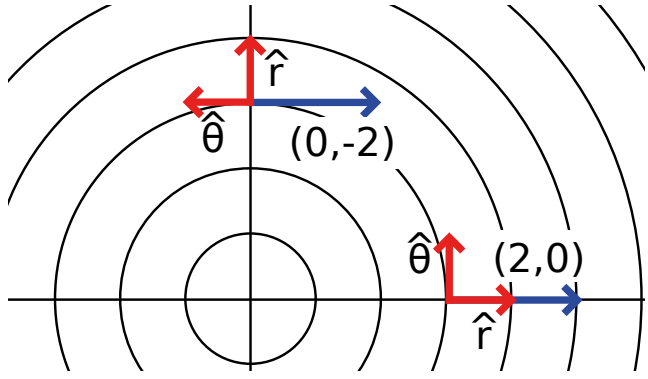

Figure 2.2. Two parallel vectors (long blue arrows) based at the points $(3, 0)$ and $(3, \frac{\pi}{2})$ in polar coordinates (r, θ). Although the two vectors are parallel, their components differ, because the tangent spaces (short red vectors) at the two points are rotated by $\frac{\pi}{2}$ relative to each other. The components of the two vectors are $(2, 0)_T$ and $(0, -2)_T$, respectively, where subscript T means that the coordinates are in the tangent spaces, as opposed to the manifold coordinates (r, θ).

$(3, \frac{\pi}{2})$ in Fig. 2.2. Although the two vectors are parallel, they do not have the same coordinates in the tangent spaces spanned by $\hat{r}$ and $\hat{\theta}$, because the two tangent planes are rotated by $\pi/2$ relative to each other. The tangent space coordinates therefore clearly have to change under parallel transport.

We can now formulate what is known as a covariant derivative

$$I^{\lambda}{}_{\nu} D_{\mu} = I^{\lambda}{}_{\nu} \partial_{\mu} + \Gamma^{\lambda}{}_{\mu\nu}, \tag{2.9}$$

where $I^{\lambda}{}_{\nu}$ is the identity matrix. Imagine applying this to a vector field of parallel vectors, in the sense of parallel transport just defined. We see that the first partial derivative may give a finite contribution, due to parallel vectors having different coordinates at different points on the manifold. What the connection does is to correct for this contribution, making the covariant derivative zero for parallel vectors. The covariant derivative is therefore in a sense closer to our usual notion of derivative than the partial derivative. If we consider what we call parallel vectors to be equal to each other, the covariant derivative preserves the notion that the derivative is zero whenever things do not change.

2.1.3 Curvature

Having understood the connection, we now move on to describe another closely related construct; the curvature. First consider parallel transport on the two surfaces in Fig. 2.3. In both cases a vector is moved along a path that finally returns the vector to its original position. For the parabolic plane, which is bent but not curved according to the terminology of differential geometry, the vector returns parallel to the original vector. However, when the vector is carried

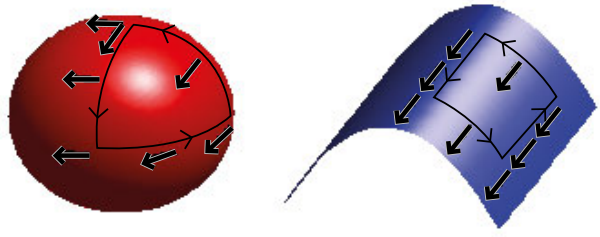

Figure 2.3. Parallel transport of a vector on a sphere and a parabolic plane. The sphere is curved, and the vector is therefore rotated as it is transported along the path. In contrast, the parabolic plane is only bent, not curved, and the vector always point in the same direction when it is returned to its original position. Note in particular, that for the sphere the vector has to be continuously re-calibrated when it is transported along the equator, in order to remain a tangent vector to the sphere.

along the path on the sphere, which is curved, the vector returns at an angle to its original direction. We emphasize that the parabolic plane is possible to arrive at by simply bending a plane, while the surface of the sphere only can be formed from a plane by stretching and compressing it. The parabolic plane is therefore intrinsically equivalent to the Euclidean plane, it is only embedded differently in three-dimensional space. In contrast, the sphere is intrinsically different from a Euclidean plane and its metric appears deformed when compared to the Euclidean plane. As defined in differential geometry, curvature therefore has to do with the non-preservation of direction of a vector when it is parallel transported around a closed path.

If we now restrict ourselves to an infinitesimal loop, we can derive an expression for how much the vector changes as it is transported around it. For simplicity we restrict ourselves to loops in the planes spanned by the basis vectors. In particular, we chose arbitrary basis indices μ and ν and denote these with square brackets $[\mu]$ and $[\nu]$ to indicate that these corresponds to specific directions rather than indices to be summed over. In Fig. 2.4, such a loop is depicted. To the first order in $\delta_{[\nu]} = |dx_{[\nu]}|$, where $dx_{[\nu]}$ is the infinitesimal vector along ν, the change in connection between nearby paths is given by $\Gamma^{\lambda}{}_{[\mu]\rho}(x + dx_{[\nu]}) = \Gamma^{\lambda}{}_{[\mu]\rho}(x) + \delta_{[\nu]}\partial_{[\nu]}\Gamma^{\lambda}{}_{[\mu]\rho}(x)$.[4] Further, we define

$$T^{\lambda}_{[\pm\mu]\rho}(x) = I^{\lambda}{}_{\rho} \mp \delta_{[\mu]}\Gamma^{\lambda}{}_{[\mu]\rho}(x), \tag{2.10}$$

where $I^{\lambda}{}_{\rho}$ is the identity. It is clear that $T^{\lambda}_{[\pm\mu]\rho}(x)$ facilitates parallel transport from x to $x \pm dx_{[\mu]}$. The same construction applies also for $\mu \leftrightarrow \nu$. The parallel transport of a vector around the loop displayed in Fig. 2.4 requires

[4] Note that no summation is implied for the index ν, which appear twice in the subscripts. The same is true for all indices that appear twice in the subscripts in this section.

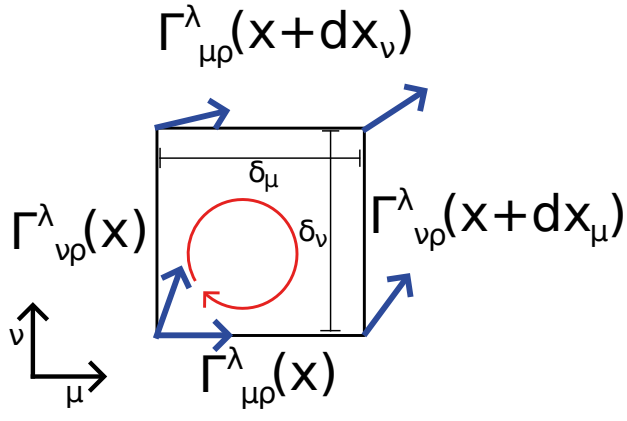

Figure 2.4. Parallel transport around a closed loop of side lengths δ_μ and δ_ν. Along each of the four paths, the parallel transport is determined by the connection along that path. When a vector is parallel transported back and forth along a single path, the end result is the same as the initial vector. However, when the vector is parallel transported along a loop, the returning vector can have a different direction. If this is the case, we call the contained area element curved.

four successive translations, which now can be written as

$$T^{\lambda}_{[-\mu]\sigma_1}(x)T^{\sigma_1}_{[-\nu]\sigma_2}(x+dx_{[\mu]})T^{\sigma_2}_{[+\mu]\sigma_3}(x+dx_{[\nu]})T^{\sigma_3}_{[+\nu]\rho}(x), \tag{2.11}$$

Expanding to first order in $\delta_{[\mu]}$ and $\delta_{[\nu]}$, and subtracting the identity, we find that the change in the vector is given by the area element $\delta_{[\mu]}\delta_{[\nu]}$ multiplied by the Riemann curvature tensor[5][6, 42]

$$R^{\lambda}{}_{\rho\mu\nu} = \partial_\mu \Gamma^{\lambda}{}_{\nu\rho} - \partial_\nu \Gamma^{\lambda}{}_{\mu\rho} + \Gamma^{\lambda}{}_{\mu\sigma}\Gamma^{\sigma}{}_{\nu\rho} - \Gamma^{\lambda}{}_{\nu\sigma}\Gamma^{\sigma}{}_{\mu\rho}. \tag{2.12}$$

The form of the Riemann curvature tensor can now be understood in the following way. To form a loop, two vectors that specify the plane in which the loop exists are needed, and these two vectors act on the μ and ν indices. Further, a vector to parallel transport is needed, which requires a third index ρ. Finally, the curvature, just like the connection, returns the change in coordinates of the returning vector, which is indexed by λ.

To arrive at the change of a vector for any finite loop, it is now in line with Stokes theorem possible to divide a surface element terminated by a loop into infinitesimal surface elements. The total change is then obtained by integrating over all these individual loops. This hints at why the integral over a curvature will turn out to be of interest next.

[5] We have here dropped the square brackets on the μ and ν indices because we initially allowed these to correspondt to arbitrarily chosen basis vectors. However, we note that strictly speaking we have here only shown that the Riemann curvature tensor is relevant for closed loops inside the planes spanned by the basis vectors, while more generally it is in fact the relevant quantity for any loop.

2.1.4 Gaussian curvature

We finally mention the Gaussian curvature by noting that in the two-dimensional case the two vectors that specify the plane in which the parallel transport is carried out are superfluous. It can be shown that in two dimensions it is enough to parametrize the curvature by a single scalar, and that the Riemann curvature tensor can be written as [39]

$$R^{\rho}{}_{\sigma\mu\nu} = G g^{\rho\lambda} \left(g_{\lambda\mu} g_{\nu\sigma} - g_{\lambda\nu} g_{\mu\sigma} \right), \tag{2.13}$$

where

$$G = \frac{1}{2} g^{\sigma\nu} R^{\rho}{}_{\sigma\rho\nu}. \tag{2.14}$$

G is called the Gaussian curvature, and when integrated over the whole surface of a sphere, it turns out that the result is [43]

$$\int G dS = 4\pi. \tag{2.15}$$

Moreover, the right hand side of the integral remains unchanged under any continuous deformation of the sphere's shape [43]. This will turn out to be a property of interest to us in the next section on topology.

2.2 Topology

2.2.1 Continuous deformations and topological equivalence classes

Topology is a field of mathematics concerned with the classification of objects into classes, such that objects are considered equivalent if they can be continuously deformed into each other [45]. The standard example is the coffee cup and the donut. Although these two objects at first may look very different, they can be continuously deformed into each other, as is indicated in Fig 2.5. In contrast, the bun has a different topology than the donut and coffee cup. Only after cutting a hole in the bun, which is a discontinuous process, can the bun be continuously deformed into the other objects. All donuts, coffee cups, and other objects that can be continuously deformed into each other are said to belong to the same equivalence class. In topology we are only interested in these equivalence classes, rather than the objects themselves.

2.2.2 Topological invariant

From a topological point of view, there are only two different objects in Fig. 2.5: the object with a hole, and the object without a hole. It is in fact the

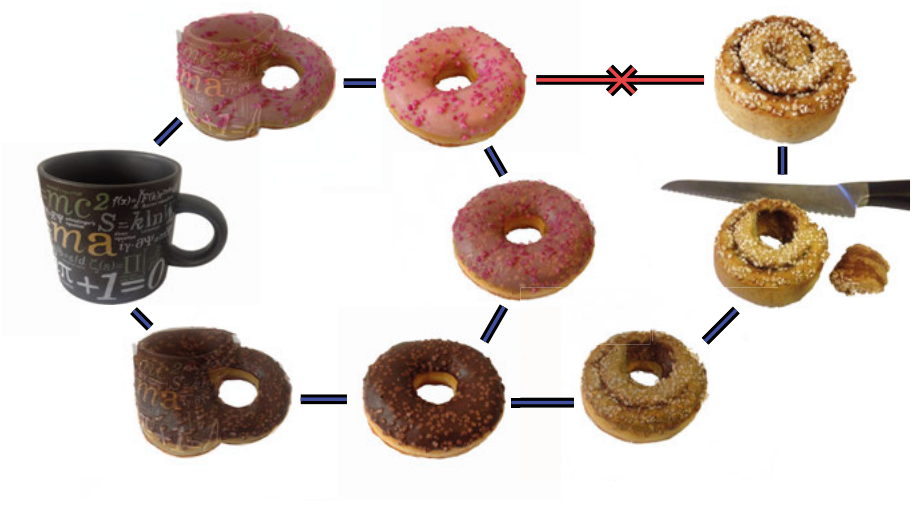

Figure 2.5. A coffee cup, a pink and a brown donut can all be continuously deformed into each other and are therefore topologically equivalent. This stands in contrast to the bun, which only becomes topologically equivalent to the others once a hole is cut in it.

case that all three-dimensional objects with a single hole can be continuously deformed into a coffee cup or a donut. In the same way all objects without a hole can be continuously deformed into the bun. These are special cases of a more general statement that any object[6] with N holes can be continuously deformed into any other object with N holes [47]. What we have encountered is an example of a topological invariant: a quantity that stays the same under continuous deformations and which can only change when we do something as drastic as cutting a hole in the bun.

2.2.3 Equivalence classes dependent on the embedding space

So far, we have only considered the objects themselves, implicitly assuming that they are embedded in an ordinary three-dimensional space. However, in general the topological classes will not only depend on the objects, but also on the spaces they are embedded in. This is demonstrated in Fig 2.6. The rightmost donut belongs to a different equivalence class, because it is threaded by a string, which prevents it from being continuously deformed into the other two. We can in this case think of the embedding space as consisting of ordinary three-dimensional space minus the points along the string. The number of holes is in this case still a topological invariant, as it stays the same under continuous deformations of the objects, and two objects that belong to the same class still necessarily have the same invariant. However, it is no longer guar-

[6] We here implicitly assume objects with a boundaryless connected and orientable surface.

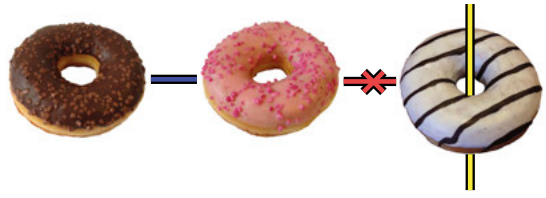

Figure 2.6. Although all three donuts give rise to the same topological invariant (number of holes), the rightmost donut belongs to a different equivalence class, because the thread prevents it from being continuously deformed into any of the other two donuts.

anteed that two objects with the same topological invariant belong to the same equivalence class. This is a general feature of topological invariants, they are often necessary but not sufficient indicators of two objects' topological equivalence [42].

2.2.4 Topological invariant as an integral over a curvature

Having understood what a topological invariant is, we consider a second example. In the previous section on differential geometry, we mentioned that the integral of the Gaussian curvature over the sphere remained invariant under continuous deformations of the sphere. In fact, it turns out that for any surface without borders, the integral is related to the number of holes h through the Euler characteristic [43]

$$\chi = \frac{1}{2\pi} \int G dS = (2 - 2h) \,. \tag{2.16}$$

This expression is useful, because it allows us to calculate the topological invariant, without having to rely on our ability to identify the number of holes from inspection.

Although it may seem simpler to count holes in a two-dimensional surface, the integral expression demonstrates a more generally applicable way of defining topological invariants. In this case, we calculate the topological invariant by taking the integral of the Gaussian curvature over the surface. In general it is common for a topological invariant to be calculated as an integral of some curvature over some manifold. This is part of a more general framework, where a topological index, which is a special type of topological invariant, is related to an integral over a characteristic class. The characteristic class is in turn derived from a curvature [6]. In this case the index is the Euler characteristic, while the characteristic class is the Gaussian curvature. In the same sense, what will be of interest to us is another topological index, the first Chern number and its relation to the first Chern class. The first Chern class will turn out to be directly related to a curvature on what is known as a complex fiber bundle and

integrating this over the manifold we arrive at the first Chern number. However, before we can continue with a description of the first Chern number, we need to understand the concept of a fiber bundle.

2.3 Fiber bundles

2.3.1 Base space and fiber

A fiber bundle is a generalization of the differential manifold. In particular, we remember from Section 2.1 that a tangent space is attached at each point of the manifold. In the language of fiber bundles, the manifold together with the set of all tangent spaces form a fiber bundle, with the manifold as base space and the tangent space as fiber. However, in contrast to the differential manifolds, where the fiber always is a Euclidean tangent space of the same dimension as the base manifold, fiber bundles are allowed to have any type of space as fiber.

Although this may sound abstract at first, a few examples can convince us that we are used to working with many types of fiber bundles in physics. Consider for example the temperature of a surface. The surface can be parametrized using two coordinates x and y, while the temperature T takes values in a third dimension. In this case, the base space can be taken to be the two-dimensional surface, while the fiber is the one-dimensional temperature space. A second familiar example is the quantum mechanical wave function. The three-dimensional space (or four-dimensional space-time) forms the base space, while the fiber is a one-dimensional complex space. More generally, if the wave function is an N-component spinor, such as in relativistic quantum mechanics where $N = 4$, the fiber instead is a four-dimensional complex vector space. All of these are examples of fiber bundles, since they share the defining feature that they consist of a base manifold to which some other space is attached at each point.

2.3.2 Connections on fiber bundles

In Riemannian geometry, the tangent spaces contain vectors that, among other things, can point us in directions within the manifold base space. Through a discussion about the embedding space, this lead us to introduce the metric and connection between different tangent spaces, or in the language of fiber bundles: connection between nearby fibers. We then learned that the embedding space is not needed at all. As long as a metric is defined on the manifold, the connection and curvature follows from the metric.

The situation is different on a fiber bundle. As we know, the fiber is not necessarily a tangent space, which means that there does not necessarily exist a metric on the manifold. In fact, when it comes to fiber bundles, we are often not foremost interested in the base manifold, or distances on that manifold. Instead we are often interested in the behavior of the fibers themselves.

Consider the case of the temperature of the surface, or the wave function in the examples above. Even though we may in these cases measure the distance between points in the surface or space, we are often mainly interested in the temperature or wave function that takes values in the fibers.

For the reasons just mentioned, we often start with a connection defined on the manifold, rather than a metric. This connection tells us how to "parallel transport" values in the fiber to nearby fibers. This may at first seem like a strange statement, because what does it mean to parallel transport something that does not necessarily have a clear geometrical interpretation. Again, a few examples can help us understand what we mean by parallel transport in this more general sense. For this purpose the example of the temperature of the surface is not the most enlightening, as the most likely way we would like to define parallel transport in this case is too trivial to teach us anything.

As a first example, let us instead consider a one-dimensional manifold that represents time. At each point of this one-dimensional manifold, we attach a one-dimensional real fiber, which can be used to indicate the median income that particular year. This is demonstrated in Fig. 2.7, where a hypothetical median income has been plotted. Now, it is not particularly useful to directly compare values from different fibers, because the income one year cannot be sensibly compared to an income another year, if we do not also know the inflation rate in between those two years. To compare two incomes, we have to transport either of them from one year's fiber, to the other year's fiber, along lines that correctly adjust for inflation. The correct connection, which sensibly determines parallel transport between fibers, should therefore in this case be determined by the inflation. Here, we have for simplicity assumed a constant inflation rate, and in Fig. 2.7 indicated two possible choices of parallel transport: one relevant which follows the inflation, and one irrelevant which preserves the value across fibers.

Continuing with the same example, we note that the most sensible definition of a constant income is not an income that has the same value in each fiber. Rather, it is more reasonable to consider an income that follows one of the exponential lines in Fig. 2.7 to be constant. Let $f(t) = I_0 e^{at}$ be a function that describes an income over the years, with a being the annual percentual increase in income. It is clear that the derivative that correctly identifies $I_0 e^{bt}$ as a constant income, where b is the inflation rate, is the covariant derivative

$$D_t = \partial_t - b. \tag{2.17}$$

That is, the covariant derivative is zero for the constant income. We note how this parallels our discussion of covariant derivative in Section 2.1.2, with $-b$ as the connection.

A second example of parallel transport in a fiber bundle, which will be more closely related to our application of it, comes from quantum mechanics itself. Here, the fiber is the complex line on which the wave function takes values. From quantum mechanics we are used to the idea that the global phase is irrel-

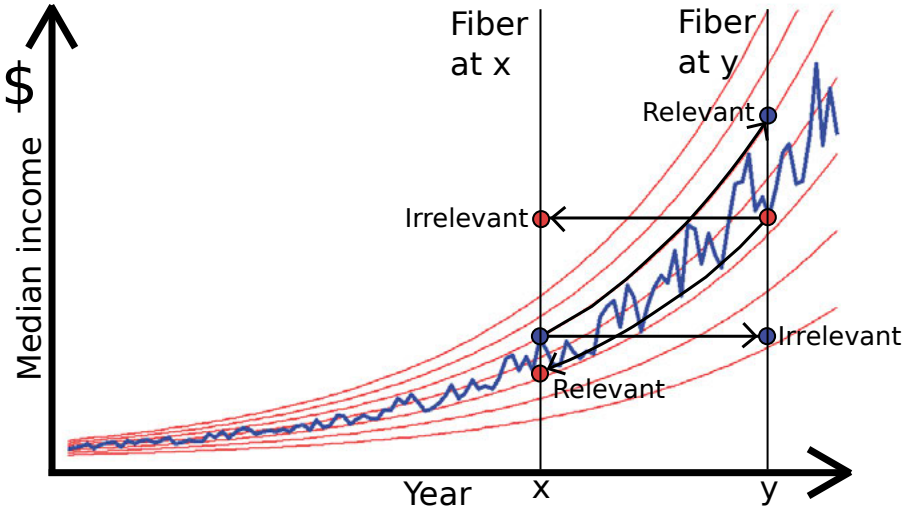

Figure 2.7. To compare earnings from different years, a direct comparison of the income is irrelevant. A more relevant comparison is obtained if we instead parallel transport the income along lines determined by the inflation and then compare them once they are on the same fiber. In this case, the most reasonable connection therefore is determined by the inflation rate. Here, a constant inflation rate has been assumed, giving simple exponential curves along which parallel transport occurs.

evant and that we are free to set this phase to whatever we want. For example, any plane wave can therefore be written as

$$\Psi(\mathbf{k}, \mathbf{x}) = Ce^{i\mathbf{k}\cdot\mathbf{x}}, \tag{2.18}$$

where C can be any complex number satisfying $|C|^2 = 1$ (we here ignore other normalization factors). In gauge theory, this freedom is promoted to each point in space, such that C becomes a function of the coordinate $\mathbf{x}$ [42]. This phase is accompanied by a corresponding change in the expression for the derivative

$$Ce^{i\mathbf{k}\cdot\mathbf{x}} \rightarrow e^{i\mathbf{k}\cdot\mathbf{x} - i\alpha(\mathbf{x})}, \tag{2.19}$$

$$\partial_\mu \rightarrow \partial_\mu + i\partial_\mu\alpha(\mathbf{x}). \tag{2.20}$$

Through this construction, it is still possible to pick down factors of k_μ from the exponent by applying the derivative to the wave function, as we are used to doing in introductory quantum mechanics

$$\left(\partial_\mu + i\partial_\mu\alpha(\mathbf{x})\right) e^{i\mathbf{k}\cdot\mathbf{x} - i\alpha(\mathbf{x})} = ik_\mu e^{i\mathbf{k}\cdot\mathbf{x} - i\alpha(\mathbf{x})}. \tag{2.21}$$

To understand what this has to do with connections and parallel transport, we now consider the fiber bundle before we introduced the local gauge transform. In this case, it is reasonable to consider elements in different fibers to be the same if they have the same value. Just like the value of an income one year ought to be compared with the value of an income another year, if it was not for inflation. That is, the function we call constant in this case is most reasonably $\Psi(0, \mathbf{x}) = C$. However, just like it is most reasonable to consider the inflation adjusted function $f(t) = e^{bt}$ constant, it is most reasonable to consider $\Psi(0, \mathbf{x}) = Ce^{-i\alpha(\mathbf{x})}$ as the constant function once we introduce the local gauge transformation. This is ensured through the definition of the covariant derivative

$$D_\mu = \partial_\mu + i\partial_\mu\alpha(\mathbf{x}). \tag{2.22}$$

The corresponding connection in this case therefore is $iA_\mu(\mathbf{x}) = i\partial_\mu\alpha(\mathbf{x})$.

Before moving on, we end this section by noting that b has zero indices, while $iA_\mu(\mathbf{k})$ has a single index. The Levi-Cevita connection in Eq. (2.8), however, has three indices. In general, a connection should indeed have three indices; one corresponding to the direction we move in, and two corresponding to the object we transport and its change, respectively. However, a one-dimensional object is transported in both of the examples above, making the two later indices superfluous. In the inflation example, we in addition only transport along one dimension, which also makes the third index superfluous.

2.3.3 Curvature on fiber bundles

Having seen how the connection can be generalized to fiber bundles, we are now ready to see how this leads to a generalized concept of curvature. To do

so, we continue our discussion using the quantum mechanical wave function as example. First, we note that in this example we consider the base manifold to be the three-dimensional space, or alternatively the four-dimensional Minkovski space, which are flat in the sense of Riemannian geometry. It is therefore clear that the curvature will have nothing to do with the curvature of the underlying base manifold. Instead, we adopt the point of view expressed in Section 2.1.3, that curvature has to do with the failure of a value in the fiber to come back to itself when it is parallel transported around a loop.

In the example involving the quantum mechanical wave function in Section 2.3.2, we arrived at a connection of the form $iA_\mu(\mathbf{x}) = i\partial_\mu\alpha(\mathbf{x})$. This particular connection, however, is flat. We can see this by picking an arbitrary value V for the wave function at an arbitrary point $\mathbf{x}(0)$. If we now move along some path $\mathbf{x}(t)$ for $t \in [0, 1]$, it is clear that with respect to the covariant derivative in Eq. (2.22), the value is kept constant if it takes on the value $Ve^{i(\alpha(\mathbf{x}(0))-\alpha(\mathbf{x}(t)))}$ at each point along the path.[7] If this path starts and ends at the same point $\mathbf{x}(0) = \mathbf{x}(1)$, we return to the same value V for any possible path, just like the vector remains unchanged when transported around a path on the bent plane in Fig. 2.3.

There is a simple reason for why the connection above leads to the conclusion that all values remain unchanged when transported around a closed loop. This is due to the fact that it is possible to define a global function $e^{-i\alpha(\mathbf{x})}$, which is constant with respect to the covariant derivative in Eq. (2.22). This, in turn, is a result of the connection being defined as $iA_\mu(\mathbf{x}) = i\partial_\mu\alpha(\mathbf{x})$. It is, however, possible to consider other connections, which cannot be written as a gradient of a scalar function $\alpha(\mathbf{x})$. Once we consider these more general connections, parallel transport can result in arbitrary changes in the phase as a value is carried around a closed loop. With the use of Stokes theorem,[8] the phase picked up when transported around such a loop can be calculated as

$$\int_S \mathcal{F} dS, \tag{2.23}$$

where S is an area element with the loop as boundary, and

$$\mathcal{F}_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{2.24}$$

Physically, we recognize Eq. (2.23) as the phase picked up by a particle encircling a magnetic flux. We also note the structural similarity between Eq. (2.24) and the Riemann curvature in Eq. (2.12), which reinforces our notion of $\mathcal{F}_{\mu\nu}$ as a kind of curvature.[9] Finally, Eq. (2.23) looks suspiciously familiar

[7] To see this, write the covariant derivative along the path as $\frac{D}{Dt} = \frac{\partial x^\mu}{\partial t} D_\mu$, such that $\frac{D}{Dt}(Ve^{i(\alpha(\mathbf{x}(0))-\alpha(\mathbf{x}(t)))}) = \frac{\partial x^\mu}{\partial t} D_\mu (Ve^{-i\alpha(\mathbf{x})})|_{\mathbf{x}=\mathbf{x}(t)} = 0$

[8] $\oint_{\partial S} A_\mu dx^\mu = \int_S (\partial_\mu A_\nu - \partial_\nu A_\mu)\, dS^{\mu\nu}$, where ∂S is the loop around which parallel transport is carried out.

[9] Note that the reason this curvature has only two indices, rather than four as in the Riemann curvature, is a result of the fiber it acts on being one-dimensional. The two other indices are

to us. In Section 2.2.4 we learned that the integral of the Gaussian curvature over the manifold is related to the Euler characteristic and the number of holes in the manifold through Eq. (2.16). This hints at the possibility that also an integral such as Eq. (2.23) can have topological meaning when S is taken to be the whole manifold. We will indeed see an example of this when we come to the Chern number.

2.3.4 Topological structure of fiber bundles

We have seen that in differential geometry the fibers are tangent planes, and vectors inside these tangent planes are directions in the base space manifold. It is therefore not surprising that the connection, which is a construct acting on the fiber rather than the manifold, still tells us something about the manifold itself. Especially, we learned in Section 2.2.4 that the integral over the Gaussian curvature, which followed from the connection, can tell us something about the topology of any two-dimensional compact surface. However, when it comes to general fiber bundles, there is no obvious relation between the base space and the fiber. Therefore, for a moment assuming that we still will be able to use the connection, curvature, and especially the integral over the curvature, to determine topological properties of our fiber bundles, a question arises: what, if not the base space itself, can it be that has a topological structure?

To answer this question, we once again use a simple example to demonstrate a more general phenomenon. This time we assume the base space to be a line segment, which can be closed to form a circle. However, before closing the circle we attach yet another line segment, say $[-1, 1]$, to each point of the first line, such that we arrive at a two dimensional strip. It is important to note that although the two dimensions are similar to each other, there is an important conceptual difference between them: we view one dimension as a base space manifold, and the other as a fiber. Further, we assume the trivial connection on the strip, which parallel transports one value in one fiber to the same value in the nearby fiber.

We now proceed to glue the base space together into a circle. When doing so, we need to identify not only the two end points of the base space with each other. We also need to identify each value in the fiber at one end point with a corresponding value in the fiber at the other end point. There are two qualitatively different ways to make this identification: in one case we glue them together by assuming that the connection transfers a value x at one end point to the value x in the fiber at the other end. Alternatively, we can choose to instead connect x in one fiber to $-x$ in the other. The difference between these two choices is demonstrated in Fig. 2.8. It is clear that in both cases the base space, the circle, is the same. The topological distinction is therefore not

therefore superfluous, as they only would take on a single value anyway. The two terms each containing two factors of the connection can be verified to disappear for the same reason.

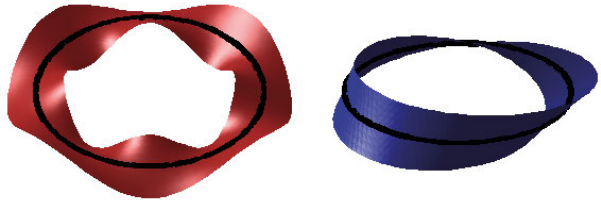

Figure 2.8. A normal strip and a Möbius strip. Both can be seen as a fiber bundle, having a circle as base space, and a line segment as fiber. Although the two fiber bundles are constructed from the same buildning blocks, they have different topology as a result of their fibers being twisted differently around the base space.

clear until we consider the whole fiber bundle. In the first case the fiber bundle is an ordinary circular strip, while the second one is a Möbius strip. This stands in contrast to our experience from differential geometry, where the topology has to do with the base space manifold alone, not the whole fiber bundle which consists of manifold plus tangent spaces.

We note the close relation between connection, topology, and parallel transport in this example. First of all, the connection at the end points is responsible for determining the topology of the whole fiber bundle, by either making it a normal strip or a Möbius strip. Secondly, if we perform parallel transport of x once around the base space, the result is x for the normal strip, while it is $-x$ for the Möbius strip. This example may seem artificial, and the division of the space into base space and fiber space may only seem to complicate things. However, it allows us to see a general phenomenon. Namely, having specified the base space and fiber, the fiber bundles constructed from these can still differ in topology. It is only once the connection is defined between all nearby fibers that the final structure of the fiber bundle is determined.

2.4 Complex vector spaces and the Chern number

We have by now covered enough background to introduce the fiber bundle and topological invariant that will be of interest to us. In particular, we are concerned with fiber bundles constructed from a base manifold together with a complex n-dimensional vector space as fiber. In the previous section we learned that the base manifold and fiber in themselves do not determine the topology of the fiber bundle. It is therefore possible to imagine that a specific choice of base manifold and a complex vector space can lead to many different topologies. This is indeed the case, and one way to classify these are through the use of Chern classes, Chern characters and Chern numbers [6, 42]. This framework is rather general, and we will therefore here limit ourselves to the case that will be of interest to us later.

2.4.1 Manifold, fiber, and a Hermitian matrix

We will be interested in Hamiltonians defined on the two-dimensional square shaped Brillouin zone and therefore choose the torus shaped Brillouin zone as base manifold. Further, as fiber we use the n-dimensional complex state vector space. Finally, we will here only describe the first Chern number in the form that is useful to us.

The fact that a Hermitian matrix, the Hamiltonian, is defined at each point of the manifold suggests that there is more structure to this problem than is implied by the manifold and fiber alone. This is indeed the case, and we will see, through explicit construction of a connection, that the existence of a Hermitian matrix suggests a natural definition for the connection. We note that the existence of a Hermitian matrix in no way forces us to choose this particular connection, rather it simply gives us a natural option for how to define one. However, once we do so, the matrix will determine the topology of the fiber bundle, and the topology of the fiber bundle will in turn contain information about the matrix.

2.4.2 Defining a connection, the Berry connection

Let $H(k_x, k_y)$ be a Hermitian matrix, which we for the moment assume to be non-degenerate and to have dimension $n \times n$. Because H is Hermitian, we know that it can be diagonalized, and the assumption that H is non-degenerate allows us to define an ordered set of n eigenvectors $|\Psi^{(\lambda)}(k_x, k_y)\rangle$ at each point (k_x, k_y). Here λ is the index that enumerates the eigenstates, and the ordering is chosen to be in terms of increasing eigenvalue. So far, there is no relation whatsoever imposed between eigenvectors at different points (k_x, k_y). However, it is natural to define parallel transport as the process whereby an eigenvector in one fiber is transported into the eigenvector in the nearby fiber which has the same index λ.

We now remember that the connection can be seen as that quantity which makes the covariant derivative zero through the relation

$$D_\mu |\Psi^\lambda\rangle = \left(\partial_\mu + iA^\rho{}_{\mu\lambda}\right)|\Psi^\lambda\rangle = 0. \tag{2.25}$$

However, as we have decided that parallel transport should occur only between states with the same label λ, $A^\rho{}_{\mu\lambda}$ has to be diagonal in λ and ρ. We denote the non-zero elements by $A^{(\lambda)}_\mu \equiv A^\lambda{}_{\mu\lambda}$, where λ now is a label rather than an index to sum over.[10] This simplifies the equation to n copies of

$$\left(\partial_\mu + iA^{(\lambda)}_\mu\right)|\Psi^{(\lambda)}\rangle = 0. \tag{2.26}$$

[10] Likewise, there is no summation implied in the right-hand side of the expression $A^{(\lambda)}_\mu \equiv A^\lambda{}_{\mu\lambda}$.

Multiplying this from the left by $\langle\Psi^{(\lambda)}|$, we arrive at

$$A_\mu^{(\lambda)}\langle\Psi^{(\lambda)}|\Psi^{(\lambda)}\rangle = i\langle\Psi^{(\lambda)}|\partial_\mu\Psi^{(\rho)}\rangle, \tag{2.27}$$

where in physics $A_\mu^{(\lambda)}$ is known as the Berry connection [48]. We further note that if we also require the eigenvectors to be normalized, then $\langle\Psi^{(\lambda)}|\partial_\mu\Psi^{(\lambda)}\rangle$ is necessarily imaginary, and the above definition of the Berry connection becomes equivalent to

$$A_\mu^{(\lambda)} = -\operatorname{Im}\left(\langle\Psi^{(\lambda)}|\partial_\mu\Psi^{(\lambda)}\rangle\right). \tag{2.28}$$

2.4.3 Berry curvature and the Chern number

Having defined a connection on the manifold, we now note that this connection is a special case of the connection for the quantum mechanical wave function in Sections 2.3.2 and 2.3.3. The only difference is that the base space now is the two-dimensional Brillouin zone instead of real space. Moreover, it is possible that the connection defined through Eq. (2.28) cannot be written as a gradient of a global scalar function $\alpha(\mathbf{k})$ [48]. Parallel transport around a closed loop can therefore give rise to the same kind of shift in phase as was discussed in Section 2.3.3. We arrive at the conclusion, that in line with Eq. (2.24), the appropriate definition of the curvature is

$$\mathcal{F}_{\mu\nu}^{(\lambda)} = \partial_\mu A_\nu^{(\lambda)} - \partial_\nu A_\mu^{(\lambda)}. \tag{2.29}$$

This is known as the Berry curvature [22], and for a finite loop S the acquired (Berry) phase is, in line with Eq. (2.23), given by

$$\int_S \mathcal{F}_{\mu\nu}^{(\lambda)} dS. \tag{2.30}$$

If the Berry curvature is multiplied by $i/2\pi$, it also becomes the simplest example of what is known as the first Chern class. Namely, the first Chern class of a fiber bundle with a one-dimensional complex fiber [6]. When this additional factor is carried over into Eq. (2.30), and S is taken to be the whole manifold, the resulting number is known as the first Chern number for the manifold and is a topological invariant for the fiber bundle [6].

We may now ask: what are the deformations under which the Chern number remains invariant? To answer this question, we revisit one of the assumptions we made earlier in this section, namely that the Hamiltonian is non-degenerate. The Chern number is in general invariant under continuous deformations of H, or alternatively $|\Psi^{(\lambda)}\rangle$, whichever we prefer to think about. However, when H becomes degenerate at some point, our construction of parallel transport breaks down. The reason is that it becomes ambiguous which state to parallel transport to, as it is impossible to order the states according to their eigenvalue.

We can therefore conclude that the Chern number is a topological invariant that is invariant under continuous deformations of H, as long as H remains non-degenerate. However, a continuous deformation that takes H from one non-degenerate state to another non-degenerate state by crossing through a degenerate state, has the possibility of changing both the topological structure of the fiber bundle and the Chern numbers calculated on it. At a transition point where the topological invariant changes, H therefore has to have a degeneracy. This observation will be tremendously useful to us in the next chapter on topological band theory and underpins the whole concept of topological band theory in the form that it is encountered in this thesis.[11]

2.4.4 Geometrical meaning

Before finishing this section, we also say a few words about the fiber bundle itself, in order to try to give a picture of what the different topological equivalence classes mean. Unfortunately, it is quite difficult to form a mental picture of this, unlike for the donut and the bun. The example of the normal strip and the Möbius strip in Section 2.3.4 does, however, provide a good starting point for a mental caricature. There, the base space manifold is a line that is joined into a circle, and the two different strips arise from different ways of identifying the fibers at the endpoint when the circle is formed. In the case we have considered here, the base space is a torus, which is a square with opposite sides identified. On top of this manifold an n-dimensional complex vector space is attached to each point as fiber. When we now perform the identification of the sides of the square to form the torus, the complex vector spaces along the edges also have to be identified. The process therefore involves gluing together the edges of a fiber bundle, with two manifold dimensions and n-complex fiber dimensions ($2 + 2n$-real dimensions). This is understandably hard to visualize, but mathematically the identification is determined by the connection, which is derived from the matrix H.

Consider now the eigenvectors that form a basis for the fiber at each point of the manifold. Whenever H is non-degenerate, we can think of this basis as rigid in the sense that the directions of the eigenvectors are fixed if they are to remain eigenvectors. Continuous deformations of H, which do not take it through a degeneracy, therefore continuously rotate this basis in such a way that the whole fiber bundle is continuously deformed. However, when a degeneracy occurs, the basis becomes floppy at the degeneracy point in the sense that at least two basis vectors can be continuously rotated into each other, without violating that the basis remains an eigenbasis. It is therefore possible to re-glue

[11] We note that it does not underpin the whole field of topological band theory in general though, as the topological invariant may be some other invariant than the Chern number for systems not considered here. However, also for other types of topological invariants, the main ideas are the same.

the fiber bundle in a different way whenever H acquires a degeneracy point, which explains why the topological structure can change when this happens.

3. Topological band theory

In the previous chapter, we introduced some of the most important mathematical structures in the fields of topology and fiber bundles. In this chapter, we will see the physical implications of these concepts. The presentation has two purposes: the first is to introduce topological band theory in a way that is possible to follow with little prior knowledge of the subject. The only requirements should be familiarity with the free electron model and band structures, as well as a basic understanding of relativistic quantum mechanics. The connection to topology at the end of the chapter also builds on the material presented in the previous chapter. The second purpose of this chapter is to introduce an appropriate terminology, which will be particularly useful for eventually generalizing these concepts to the treatment of topological superconductors and Majorana fermions. The reader is in particular referred to the following sources for further reading [5, 20–22].

3.1 Hybridization and band theory

3.1.1 Nearly free electron model

We begin by reminding ourselves about hybridization through a simple example. Two energy levels split by an energy difference of $2\tilde{E}$, and coupled by an interaction of strength Λ, can be described by the Hamiltonian

$$H = \begin{bmatrix} \tilde{E} & \Lambda \\ \Lambda^* & -\tilde{E} \end{bmatrix}. \tag{3.1}$$

Diagonalizing this, we find that the energies are given by $E = \pm\sqrt{|\Lambda|^2 + \tilde{E}^2}$. That is, a coupling between the two states hybridizes them, and pushes their energy apart from each other. When the energy split $2\tilde{E}$ is small, the hybridization is strong, and is in the limit of complete degeneracy ($\tilde{E} = 0$) pushing apart the two levels by $2|\Lambda|$. On the other hand, when $\tilde{E}$ is large, the hybridization energy is negligible. This means that any coupling that occurs between quantum mechanical energy levels will tend to split them, and especially so for degenerate ones.

This simple observation has major implications in condensed matter theory, where band gaps often arise for exactly this reason. In the nearly free electron model for example, energy levels are initially assumed to have a k^2-dispersion.

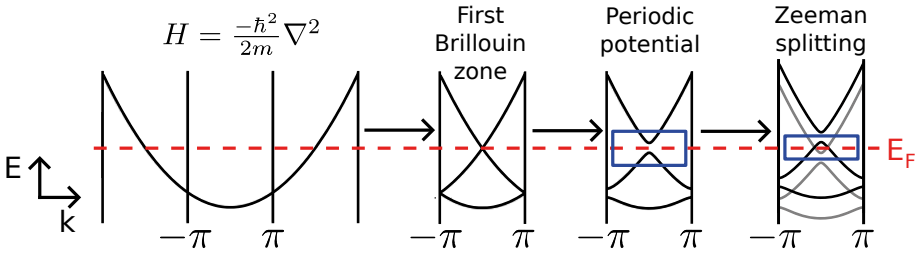

Figure 3.1. In the nearly free electron model, the starting point is a single infinite parabolic band. By dividing k-space into Brillouin zones and translating all branches of the similarly divided band into the first Brillouin zone, a series of overlapping bands are stacked on top of each other. Introducing a periodic potential, which couples vertically aligned energy levels, gaps open up at degeneracy points. The result is multiple bands, separated by gaps, which to a first approximation have a parabolic dispersion around the points where gaps have opened up. A horizontal dashed Fermi level line has been inserted to demonstrate that such band splittings can give rise to an effective low energy theory in the left blue box. Such theories are to second order described by two parabolas that bend away in opposite directions and are separated by a gap. Finally, two such parabolas can be made to overlap by introducing a Zeeman term, as seen in the second blue box.

Due to the periodicity of the lattice, k-points separated by reciprocal lattice vectors couple to each other, and split the bands around the Brillouin zone boundary and the Γ-point ($\mathbf{k} = 0$). Such splitting gives rise to bands that assume parabolic shapes around the split point, with parabolas bending both upward and downward. See the first three steps in Fig. 3.1. [49, 50]

3.1.2 Parabolic bands

Most problems in condensed matter physics cannot be solved exactly. Rather, we often need to attack it through a series of approximations, step by step adding more terms to the Hamiltonian, to incrementally capture more details. It is therefore not an uncommon scenario to in one step have arrived at a dispersion relation, which to a first approximation can be seen as two parabolas, each bending in opposite direction and being separated by a band gap $2M$. At the next level of approximation it is then also possible that these two bands couple to each other by yet another term. Such a system is described by the Hamiltonian in Eq. (3.1), if we replace $\tilde{E} \to M + |\mathbf{k}|^2$. The energy is in this case given by

$$E = \pm \sqrt{|\Lambda|^2 + (M + |\mathbf{k}|^2)^2}. \tag{3.2}$$

In the fourth step in Fig. 3.1, we see how the addition of a Zeeman term to the nearly free electron model can result in a low energy theory of this form. M

is in this case related to the original band gap in step three, and the strength of the Zeeman term. A coupling Λ would further be provided through any term that can connect the two spin species to each other. This is, however, only an example of how to arrive at such a theory. To keep the discussion general, we will from here on only assume that we for some reason have two parabolic bands with the dispersion relation in Eq. (3.2). The underlying physical reason for the existence of such bands will be left open for specification in each particular physical realization of interest.

3.1.3 Band inversion and Rashba interaction

In the example above, we considered the coupling Λ to be the same at all k-points. Often this simplification is not true, and in some cases, such as for the Rashba spin-orbit interaction, the coupling is even guaranteed to be exactly zero at some points. The Rashba spin-orbit interaction is a term that can be derived from relativistic quantum mechanics, which arise as a spinful particle move through a uniaxial electric potential, and has the form [51]

$$H_{\text{Rashba}} = \alpha \left(\boldsymbol{\sigma} \times \mathbf{k}\right)_z = \begin{bmatrix} 0 & \alpha(k_y + ik_x) \\ \alpha(k_y - ik_x) & 0 \end{bmatrix}. \tag{3.3}$$

Now assuming a two-dimensional band structure, such that $|\mathbf{k}|^2 = k_x^2 + k_y^2$, we see that a Rashba-like coupling between two bands bending away from each other can be described by setting $\Lambda = \alpha(k_y + ik_x)$. This means that Eq. (3.1) and (3.2) become

$$H = \begin{bmatrix} M + k_x^2 + k_y^2 & \alpha(k_y + ik_x) \\ \alpha(k_y - ik_x) & -M - k_x^2 - k_y^2 \end{bmatrix}, \tag{3.4}$$

$$E = \pm \sqrt{\alpha^2(k_x^2 + k_y^2) + (M + k_x^2 + k_y^2)^2}. \tag{3.5}$$

In Fig. 3.2, this dispersion relation has been plotted along one direction in k-space for different M. We first note that for the red curves, which correspond to the Rashba coupling being turned off, the bands go from a non-overlapping to an overlapping regime as M is tuned from positive to negative. This we call going from a normal band order to an inverted band order.[1]

Once the Rashba term is turned on, the band structure with $M > 0$ remains fairly unchanged, due to the fact that the coupling term acts on states already separated by a sizeable energy. However, in the inverted regime $M < 0$ the situation is very different. There, the Rashba term can couple the states that originally were degenerate and reopen a gap in the previously gapless spectrum. Finally, we note that in the intermediate case $M = 0$, the two bands

[1] Note that band inversion refers to the energy of the point $\mathbf{k} = 0$ with no kinetic energy. That is, the bands are called inverted if the potential energies have crossed each other.

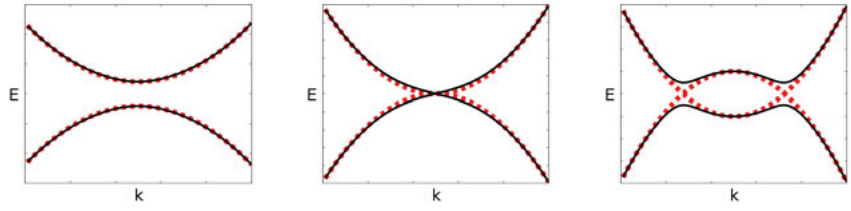

Figure 3.2. Band structure given by the dispersion relation in Eq. (3.5), plotted along the line $k_y = 0$. From left to right $M > 0$, $M = 0$, and $M < 0$. (Red dotted) Two parabolic bands progressing from normal to inverted band order. (Black) Same as red, but with an additional Rashba like coupling $\alpha \neq 0$ between the two bands. At $M = 0$, the Rashba term fails to split the two bands, because the coupling strength is exactly zero at $\mathbf{k} = 0$.

touch at $\mathbf{k} = 0$, but the Rashba term is unable to open up a gap because it is zero at this point. In this special case, the band structure therefore remains gapless, even when the coupling between the two bands is turned on.

3.2 Dirac cone and negative mass gap

Let us now study the special intermediate case $M = 0$ closer. In this case the dispersion relation reduces to

$$E_{M=0} = \pm\sqrt{\alpha^2(k_x^2 + k_y^2) + (k_x^2 + k_y^2)^2} \approx \pm\sqrt{\alpha^2(k_x^2 + k_y^2)} = \pm\alpha|\mathbf{k}|. \tag{3.6}$$

In the second step, it is assumed that we are only interested in points close to $\mathbf{k} = 0$, where the two bands meet. There, the dispersion relation is linear in $\mathbf{k}$, which means that close to $\mathbf{k} = 0$, the bands form a cone. This is known as a Dirac cone, which can be understood by studying the case also for a small but finite M. Let us for simplicity assume that $|M| \ll \frac{\alpha^2}{2}$, as this means that the gap is still smallest at $\mathbf{k} = 0$. We can then once again approximate the bands around $\mathbf{k} = 0$ to obtain

$$E \approx \pm\sqrt{M^2 + \tilde{\alpha}^2(k_x^2 + k_y^2)}, \tag{3.7}$$

where $\tilde{\alpha}^2 = \alpha^2 + 2M \approx \alpha^2$. Comparing this to the relativistic dispersion relation satisfied by Dirac fermions[2] [52, 53]

$$E_{\text{Rel}} = \pm\sqrt{m^2 + |\mathbf{k}|^2}, \tag{3.8}$$

where m is the particle mass, it is possible to make a formal analogy between the two dispersion relations. For the sake of the analogy, we also assume that

[2] Natural units assumed, where $c = 1$.

the Fermi level is situated in the middle of the gap (at the crossing point when $M = 0$). Then the holes in the valence band and the electrons in the conduction band become analogous to the positrons and electrons in the Dirac equation [52, 53]. This is the reason why the dispersion relation in Eq. (3.7) is called a Dirac spectrum, and why the special case at $M = 0$ is called a Dirac cone [54]. Now, because M plays the role of a mass, we also refer to the gap as a positive or a negative mass gap, depending on the sign of M.

3.3 Bulk-boundary correspondence

In the previous section, we saw how fine-tuning of M can result in a spectrum with a cone structure around $\mathbf{k} = 0$. Now, it is hard to imagine that such fine tuning can actually occur in a realistic material. Especially once we consider that in realistic materials different types of impurities and dislocations can introduce coupling parameters into the problem that are hard to control. There is, however, an alternative to fine-tuning by which we can be guaranteed to obtain a Dirac cone, but the price we have to pay is one dimension.

To understand what is meant by this, we imagine a material that on one side has been tuned deep into the $M > 0$ regime, while the other side is tuned deep into the $M < 0$ regime. This requires no fine-tuning as we are not interested in the exact value of M. We even allow for M to vary inside the regions, as long as it does not change sign, rendering impurities and dislocations that introduce couplings of size smaller than the gap irrelevant. As we move from one side of the sample to the other, somewhere along this path there needs to be an interface on which M changes sign. Along this interface, we therefore expect to once again have a gapless Dirac cone. This cone will, however, have one dimension less than the cone in the previously imagined fine tuned bulk, as the interface has one dimension less than the bulk. A schematic picture of this argument can be seen in Fig. 3.3.

In the argument above, we have considered a single material, which for some reason can be tuned into regimes with different sign of the mass gap. While it is possible to imagine that this can be achieved in reality, by e.g. doping the two sides differently, we can simplify the setup further. Assume therefore that we keep turning up M on the positive mass gap side. Eventually the gap will be so big that there is no possibility to excite a particle at all. This state is very similar to a vacuum, where an enormous amount of energy is required to excite (create) a particle. We can therefore replace the positive mass gap material by the vacuum in the argument above [20]. It follows that a material with a negative mass gap, which is surrounded by vacuum, is guaranteed to have gapless edge states on its boundary. On the other hand, a material that is in the positive mass gap regime does not need to have gapless boundary modes. The reason is that there is no need for a change of sign of M as we go from a small positive M inside the material to a large one in the vac-

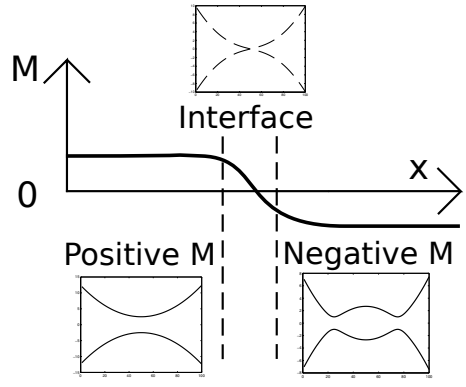

Figure 3.3. Two different sides of a material are tuned deep into the $M > 0$ and $M < 0$ regimes, respectively. Somewhere in between these two sides an interface must exist where M crosses through zero. At this interface, the spectrum is expected to regain the form of a Dirac cone, although the cone now exists in one dimension less compared to the bulk.

uum outside. This is an example of the bulk-boundary correspondence [20]: a parameter in the bulk, in this case the sign of M, which only changes as the bulk passes through a band closing, can tell us something about the boundary of the system. In this case, gapless edge states are guaranteed to exist on the boundary, if the mass gap in the bulk is negative.

3.4 Edge states

3.4.1 Localization of a half Dirac cone per edge

The above discussion is qualitatively correct. However, we note that the argument implicitly assumes that it is possible to talk about real space and momentum space at the same time. This is no problem as long as M varies slowly in space,[3] but in reality we expect interfaces to be abrupt regions of size no more than a few lattice constants. A little more detailed analysis is therefore needed to justify the above discussion also in this case.

To do so, we assume an interface at $y = 0$, separating the two regions $y < 0$ and $y > 0$, for which the mass gap is given by $M(y) = M\text{sgn}(y)$. Focusing on the low energy part of the Hamiltonian in Eq. (3.4), where $k_x^2, k_y^2 \to 0$, and

[3] When M varies slowly in space, we can divide space into cubes small enough that M can be considered constant inside each volume element, but large enough that we can treat the region with a Fourier series with k-point spacing close enough to be considered continuous.

performing an inverse Fourier transform along the y-direction, we arrive at

$$H = \begin{bmatrix} M(y) & i\alpha(k_x - \partial_y) \\ -i\alpha(k_x + \partial_y) & -M(y) \end{bmatrix}. \tag{3.9}$$

Separating the Hamiltonian as $H = H_x + H_y$, we see that the equation in the x-direction has a vanishing mass gap

$$H_x = \begin{bmatrix} 0 & i\alpha k_x \\ -i\alpha k_x & 0 \end{bmatrix}. \tag{3.10}$$

However, a solution to this Hamiltonian needs to simultaneously be an eigenstate of

$$H_y = \begin{bmatrix} M(y) & -i\alpha\partial_y \\ -i\alpha\partial_y & -M(y) \end{bmatrix}, \tag{3.11}$$

in order to be an eigenstate of the total Hamiltonian. Moreover, in order for the total dispersion relation to be given by H_x, it is important that the energy contribution from H_y is zero. Therefore, we further require the eigenstate to be an eigenstate with zero energy contribution from the partial eigenvalue problem $H_y|\Psi\rangle = 0$. Such a solution is indeed provided by [20, 21]

$$|\Psi^{(-)}_{k_x}\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix} e^{-\int_0^y \frac{M(y')}{\alpha} dy'}. \tag{3.12}$$

To understand this eigenstate, we note that the integral in the exponent is a positive number. This is so, because if $y > 0$ we integrate forward with a positive integrand, while if $y < 0$ the integration is backwards over a negative integrand. The exponent is therefore decaying away from the interface, which means that the state is bound to the surface. Applying this state to H_x, we also see that the dispersion relation is $E(k_x) = -\alpha k_x$, which is one half of a Dirac cone. We further note that it is useful to think of the division of H into two pieces as a division of the energy into dispersive and localization energy. The dispersive energy is contributed by H_x, while the localization energy is provided by H_y. The edge states are therefore states that cost zero energy to localize on the edge.

Now consider the eigenstate

$$|\Psi^{(+)}_{k_x}\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix} e^{\int_0^y \frac{M(y')}{\alpha} dy'}, \tag{3.13}$$

which at first seems to be a valid eigenstate as well, with dispersion relation $E(k_x) = \alpha k_x$. This is the second half of the Dirac cone. However, this is in fact not a physically valid eigenstate, because the exponent here is increasing away from the interface. It becomes infinitely large inside the material. This means that as we squeeze the width over which the transition in $M(y)$

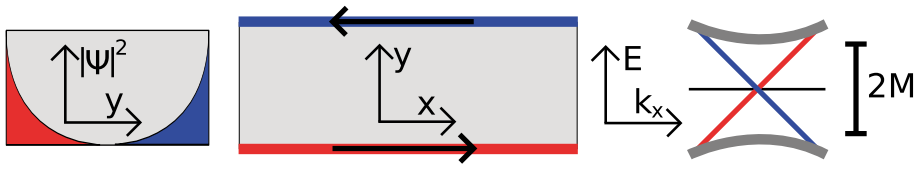

Figure 3.4. Edge state dispersion along the x-axis for an infinite strip. (Left) The right and left moving branches are exponentially localized on different edges of the infinite strip. (Center) The infinite strip in real space. The edge channels propagate current in opposite directions along the upper and lower edge. (Right) Total dispersion relation. Two bulk bands with mass gap M are shown in grey, while the right and left moving edge states together form a Dirac cone. The two branches of the Dirac cone are localized at opposite edges of the strip.

is occurring, only that branch of the Dirac cone that has negative slope survives. Alternatively, we can say that only one of the two branches can satisfy a boundary condition that localizes it on the interface. The other solution is, however, physically meaningful if we instead take $M(y) \to M(-y)$. This means that the other branch will localize on an edge, where the change in mass gap happens in the opposite direction. If we therefore instead define $M(y)$ by

$$M(y) = \begin{cases} -M & -Y < y < Y \\ M & \text{otherwise} \end{cases}, \tag{3.14}$$

the first branch becomes localized at $y = Y$, while the second branch localizes at $y = -Y$. See Fig. 3.4 for a conceptual picture.

3.4.2 HgTe quantum wells

Returning to the case $M(y) = M\text{sgn}(y)$ and a single edge, we have seen that the spin that propagates along the edge with dispersion $E(k_x) = -\alpha k_x$ is $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & i \end{bmatrix}^T$. Had the Hamiltonian, however, been given by $H^*(-\mathbf{k})$, we would have arrived at the same conclusion with the dispersion relation $E(k_x) = \alpha k_x$ and spin state $\begin{bmatrix} 1 & -i \end{bmatrix}^T$. This is for example the case for some topological insulators, where the first model predicting that a HgTe-CdTe quantum well is a two-dimensional topological insulator involves both of these two Hamiltonians.

In the model of HgTe, the Rashba like coupling is acting on a pseudo-spin, while the two spin-species are described by the two Hamiltonians $H(\mathbf{k})$ and $H^*(-\mathbf{k})$ [14]. $H(\mathbf{k})$ and $H^*(-\mathbf{k})$ are therefore both two-by-two matrices written in a pseudo-spin basis. It is clear that the Hamiltonian that describes the total system therefore can be formulated through the use of the four-by-four

block-diagonal matrix

$$\mathcal{H}(\mathbf{k}) = \begin{bmatrix} H(\mathbf{k}) & 0 \\ 0 & H^*(-\mathbf{k}) \end{bmatrix}. \tag{3.15}$$

Here each block gives rise to its own edge state, and because of the different dispersion relations $E(k_x) = -\alpha k_x$ and $E(k_x) = \alpha k_x$ for the two blocks, the two edge states are counter propagating. Further, because each block describes a different spin-species, the two branches propagates different types of spins. In topological insulators the edge states therefore consists of two counter propagating channels of opposite spin, so called helical edge states.

3.4.3 Robustness of the edge states

Having seen how the bulk Hamiltonian is related to the dispersion on the edge, we now turn to a more detailed discussion of impurities. For this purpose we assume that the bulk is in the negative mass gap regime, such that the bulk-boundary correspondence predicts gapless edge states to exist. We already mentioned above that impurities in the bulk are of little importance, as long as they are small enough to not change the sign of the mass gap. But what about impurities close to, or at, the boundary? We will here see that it is possible to classify these into irrelevant and relevant impurities. Irrelevant impurities are those impurities that just like impurities in the bulk does not destroy the gapless edge spectrum. Relevant impurities are on the other hand those that possibly can open up a gap on the edge.

We begin with the trivially irrelevant impurities, which perturb the diagonal entries of the Hamiltonian. As they only affect the diagonal entries, we can think of these impurities as local disturbances of the mass gap parameter. In the bulk we already know that such impurities are irrelevant, because the mass gap is large there. However, close to the edge the mass gap becomes small, and the impurity can therefore locally deform the mass parameter into the opposite regime. In spite of this, these impurities are unable to destroy the existence of gapless edge states. The reason is that although they locally can deform the mass gap parameter, they do not change the fact that the mass gap has to change sign somewhere in between the bulk and the vacuum. All these impurities can do is therefore to deform the shape of the interface, not remove it [21].

Next, for simplicitly we assume that the interface is along the x-direction, such that the Hamiltonian along the edge is given by Eq. (3.10). As we know that diagonal entries are irrelevant, we now try to open up a gap by hybridizing states through the use of off-diagonal elements. It is however impossible to couple states in this Hamiltonian, because although it appears to have Kramer's degenerate pairs at k_x and $-k_x$, we learned above that only one of these branches localizes on the interface. Further, we know that the absence of

degenerate states means negligible hybridization.[4] It therefore seems like all impurities are irrelevant.

We now note, however, that the topological insulator mentioned in the section above in fact consisted of two copies of the Hamiltonian considered so far. Each copy gives rise to its own branch, which is both counter propagating and has opposite spin relative to the other branch. This means that the argument in the previous paragraph falls apart for this system. There are now Kramer's degenerate pairs at k_x and $-k_x$, which indeed can couple to each other. The coupling only needs to couple up spins to down spins, and an impurity that introduces a properly aligned time reversal symmetry breaking Zeeman term therefore suffices in this case [21]. Relevant impurities are therefore a reality, but they do need to satisfy certain conditions, making many of the impurities we usually need to care about irrelevant to the edge states.

3.5 Lattice model

So far we have mentioned the Brillouin zone, but have not made explicit use of it. Rather, we have considered the problem using a continuum model, where **k** in principle can take values in the whole real plane. Here we reformulate the problem to a lattice model, by making the replacement $k_\mu \to \sin(k_\mu)$ and $k_\mu^2 \to 1 - \cos(k_\mu)$. The Hamiltonian in Eq. (3.4) then becomes

$$H = \begin{bmatrix} d_z & d_x - id_y \\ d_x + id_y & -d_z \end{bmatrix}, \tag{3.16}$$

where

$$d_x = \alpha \sin(k_x), \tag{3.17}$$
$$d_y = \alpha \sin(k_y), \tag{3.18}$$
$$d_z = M + 2 - \cos(k_x) - \cos(k_y). \tag{3.19}$$

For small **k**, we see that the continuum Hamiltonian in Eq. (3.4) is recovered. The difference is in the behavior at large **k**. In the continuum model the bands go of to infinity, while in the lattice model the bands flatten out again around the Brillouin zone boundaries. As can be seen in Fig. 3.5, the band structure is essentially the same as that for the continuum model, as long as $M \gtrsim -1$. However, while in the continuum model the two $\alpha = 0$ bands overlap for all $M < 0$, the situation is different in the lattice model. At $M = -4$, the band structure once again pass into a regime of non-overlapping bands, and it turns

[4] It is indeed possible to consider couplings between nearby k-points as a possible way around this argument. In order to couple points that are close to each other within the Brillouin zone, however, requires a scattering potential that has a periodicity that is orders of magnitude larger than the lattice constant.

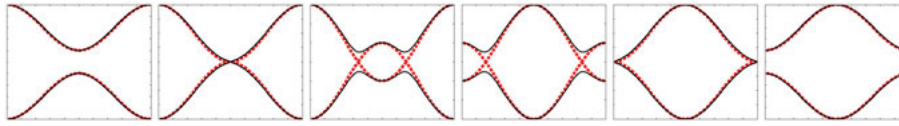

Figure 3.5. The dispersion relation along $k_y = 0$ which follows from the lattice model in Eq. (3.16) for six values of M. For $M \gtrsim -1$ the dispersion relation is qualitatively similar to the dispersion relation for the continuum model shown in Fig. 3.2. However, for $M < -4$, the two (red) $\alpha = 0$ bands become non-overlapping again. This results in only the regime $-4 < M < 0$ being guaranteed to host gapless edge states. This stands in contrast to the continuum model for which the regime with gapless states are $M < 0$.

out that now also the limit $M \to -\infty$ is similar to the vacuum [22]. This means that the only regime in which we now are guaranteed to have gapless edge states is when $-4 < M < 0$.

3.6 Topological invariant

The reader may by now have noted a few suggestive similarities between this and the previous chapter. First of all, we note the importance the sign of the mass gap has played so far. This is clearly a quantity that is invariant under continuous changes of M, but which changes discontinuously at $M = 0$. In this sense it is a topological invariant, which distinguishes between two equivalence classes of band structures; the band structures with positive and negative mass gap, respectively. The equivalence of cutting a hole in the bun here occurs when the two bands meet at $M = 0$.

Second, we note that the lattice Hamiltonian is a Hermitian matrix defined on a two-dimensional torus, and the eigenvectors of this Hamiltonian takes values in a complex vector space. We can therefore think of the problem as a two-dimensional compact manifold, with a two-dimensional complex vector space attached to each point as a fiber. The Brillouin zone and complex vector space together form a fiber bundle. We also remember from Section 2.4, that when a non-degenerate Hermitian matrix is defined on the base manifold, it is possible to define a connection that facilitates parallel transport on the fiber bundle. This in turn defines a topology of the fiber bundle. The requirement in Section 2.4.2 that the matrix is non-degenerate corresponds to the requirement that the two bands in our model above do not touch. Further, the index λ used to order eigenvectors according to their eigenvalues corresponds to the band index.

The notes above suggest that the first Chern number might be a relevant topological invariant for the band structures studied so far, and this is indeed the case. Moreover, it can be shown that the first Chern number in Eq. (2.30)

for the lower band, which results from carrying out the integral over the whole plane in the continuum case, is directly related to the sign of the mass gap through [22]

$$I^{(-)} = \frac{1}{2\pi} \int \mathcal{F}^{(-)}_{\mu\nu} dk_x dk_y = \frac{\text{sgn}(M)}{2}. \tag{3.20}$$

Here $(-)$ is the band index λ of the lower band. Note the striking similarity between this expression and the Euler characteristic in Eq. (2.16). There we saw that the Euler characteristic can be expressed either as an integral of the Gaussian curvature over the whole manifold, or as an algebraic expression involving the number of holes. That is, there is a simple relation between the Gaussian curvature, and the number of holes. In the same way we here see that the first Chern number, expressed as an integral of the Berry curvature, is directly related to the sign of the mass gap. So far the sign of the mass gap has been possible to extract manually from the model, by identifying the two parabolas and see if they overlap or not. However, in general the Hamiltonian and band structure is not as clean as the model considered so far. The first Chern number therefore has an advantage, because it is possible to calculate for a more general Hamiltonian. The integral expression provides a general way for calculating topological invariants that can tell us something about the edge states. This is an example of the framework of topological indices mentioned in Section 2.2.4, whereby the curvature is not only related to the index, but also to the modes bound to the manifold's boundary [6].

The power of the first Chern number becomes even clearer once we realize that the analytical expression for the connection is not required. As was originally shown for the quantum Hall system, the first Chern number can be calculated for any band λ using the alternative expression [10, 22, 48]

$$I^{(\lambda)} = -\frac{1}{2\pi} \int dk_x dk_y \text{Im} \left(\sum_{\rho \neq \lambda} \epsilon^{ij} \frac{\langle \Psi^{(\lambda)} | \partial_{k_i} H | \Psi^{(\rho)} \rangle \langle \Psi^{(\rho)} | \partial_{k_j} H | \Psi^{(\lambda)} \rangle}{(E^{(\rho)} - E^{(\lambda)})^2} \right). \tag{3.21}$$

The advantage of this expression is that it uses the Hamiltonian directly, instead of going through the construction of the connection. It is also insensitive to the choice of gauge for the eigenstates [22]. This lends itself for numerical calculations, because numerically eigenstates at different (k_x, k_y) are calculated independently of each other. This leads to an independent, potentially random and discontinuous gauge, making the derivative $|\partial_{k_i} \Psi^{(\lambda)}\rangle$ in Eq. (2.28) difficult to handle.

Let us now consider the lattice model instead. In this case the relation between the first Chern number and the mass gap parameter turns out to be slightly modified from that in Eq. (3.20). In the two regimes $M > 0$ and $M < -4$, the first Chern number is zero, indicating that these are trivial phases

without gapless edge states. In the overlapping regime $-4 < M < 0$, however, the first Chern number takes on two different values depending on if M is larger or smaller than -2. For $-2 < M < 0$ the Chern number is -1, while it is 1 for $-4 < M < -2$ [22]. First of all, this means that both of the two later regimes, as expected, are non-trivial. Second, because the Chern number changes at $M = -2$, we can conclude that the Hamiltonian need to be degenerate at this point. We can indeed confirm this by diagonalizing Eq. (3.16) for $M = -2$, to obtain the dispersion relation

$$E_{M=-2} = \pm\sqrt{2(1 + \cos(k_x)\cos(k_y))}. \tag{3.22}$$

There are clearly degeneracies at both $(0, \pi)$ and $(\pi, 0)$. This demonstrates how knowledge about the first Chern number can lead to conclusions about the dispersion relation, which are not entirely obvious from Eq. (3.16).

3.7 On the generalization to n-by-n Hamiltonians

We have seen that the combination of a band hybridizing term that is zero at certain points in the Brillouin zone, together with a mass gap parameter that can change sign, can have important implications for the boundary spectrum of a system. We have also seen that the first Chern number can be used as a tool for determining these properties from the bulk Hamiltonian. The demonstration of this has here been carried out in the special case of a two-by-two Hamiltonian. However, from Section 2.4 in the previous chapter, we know that the first Chern number can be defined also for band structures involving more than two bands. We can therefore expect similar features to occur also in more general cases, and indeed, we will later see that this is the case when we apply these ideas to what is called a topological superconductor. In fact, we have already touched upon an example of a four-by-four case, the topological insulator that was mentioned in Section 3.4. We saw that in this case we do not get a single, but two counter propagating helical edge states.

3.8 A note on $\mathbb{Z}_2$ classification

Although the properties of the model of HgTe are related to a Chern number, we note that this model is not a perfect example of a system with a non-trivial Chern number. In Section 2.4 we defined the Chern number for each band, using that H is a non-degenerate matrix. The model for HgTe in Eq. (3.15) is on the other hand degenerate. This means that although the two-by-two submatrices of the HgTe Hamiltonian have well defined Chern numbers, it is not clear how to define Chern numbers for the whole Hamiltonian. In spite of this, it turns out that even in the presence of degeneracies, it is possible to define

a related topological invariant. The only requirement is that no band crosses the Fermi level, and the invariant is the sum of the Chern numbers of all bands below the Fermi level [20]. However, the Chern numbers for the two different blocks in Eq. (3.15) are related by a minus sign. Their sum will therefore always be zero, and such an invariant is therefore not of much use. This is the case for any Hamiltonian with time reversal symmetry [20]. The difference between the Chern number for the occupied bands, however, is also useful as long as the matrix is block diagonal, and can be used as a topological invariant for the whole system. Dividing this by two, and only considering whether the result is even or odd, this becomes another topological invariant known as a $\mathbb{Z}_2$ invariant [20, 55]. This construction, however, is only possible as long as the Hamiltonian is block diagonal as in Eq. (3.15), while the $\mathbb{Z}_2$ invariant can be defined more generally [55]. In general, there are therefore Hamiltonians with topologically non-trivial behavior, which are not possible to classify using a Chern number. However, in this thesis we restrict our treatment to those that can, which in particular requires time-reversal symmetry breaking.

4. Superconductivity

In the two previous chapters, we have learned about topology and how it can be applied to band theory to make robust predictions about materials edges from their bulk properties. We now switch gear and introduce superconductivity, which will prepare us for a discussion of topological superconductors in the next chapter.

Superconductivity is in its simplest formulation a phenomenon whereby, at low temperatures, an ordinary conductor suddenly becomes a perfect conductor that conducts electricity without resistance. Superconducting materials would therefore be ideal for power lines, electronics, and many electric devices where resistance gives rise to unwanted energy losses, if it were not for the low temperatures required. In contrast, it is not wanted in a traditional light bulb, which relies on resistance for heating the wire that shines at high temperatures. These implications are at least on a conceptual level accessible to anyone with even the most rudimentary understanding of the technology our society is built on.

As deceptively simple as the consequences of superconductivity may seem, as complicated can it be to understand what the theory that explains its causes actually means. Many aspects of the theory that underpins our understanding of superconductivity are notoriously inaccessible to anyone who encounters the field for the first time. Being a fundamentally quantum mechanical phenomenon, it is no surprise that intuition of the underlying physics often is hard won. Explanations that try to shine light on the underlying physics often suffer from being either superficial or impenetrably technical, or both.

We will not pretend to do better here, but will in what follows attempt at giving a coherent introduction to those concepts in superconductivity that are important for understanding the mean-field Bogoliubov-de Gennes treatment of the subject. In particular, this introduction is written with the intention of introducing a terminology, which in combination with that introduced in the previous chapters, makes it possible to understand the subject of topological superconductors and Majorana fermions. At points the presentation therefore advocates a point of view that semantically may appear slightly unconventional compared to other introductions to superconductivity. For further reading on superconductivity in general, we refer to the references [56–59].

4.1 BCS theory

Although superconductivity as a phenomenon has been known since 1911, when it was first observed by Heike Kamerlingh Onnes [60], the explanation

of its microscopic origin had to wait until 1957, when Bardeen, Cooper, and Schrieffer (BCS) formulated the first truly successful microscopic theory [56]. Known as BCS theory, it has been tremendously successful at describing a large class of superconductors today known as conventional superconductors. In addition, many of its concepts continue to be important building blocks also in theories for the so far less understood unconventional superconductors [61].

4.1.1 Cooper pairs and the BCS Hamiltonian

To begin with, conventional superconductivity is a result of a many body interaction between electrons and the lattice [56]. Just like in the nearly free electron model, we can for simplicity imagine electrons in a homogeneous background. The background is assumed to successfully screen out the interactions, such that the resulting quasi-particles, which we will call electrons, can be treated just as normal electrons described by the Hamiltonian

$$\mathcal{H} = \sum_{\mathbf{k}\sigma\sigma'} H_{\sigma\sigma'}(\mathbf{k}) c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma'}. \tag{4.1}$$

Here the Hamiltonian has been written down using the language of second quantization, where $c^\dagger_{\mathbf{k}\sigma}$ and $c_{\mathbf{k}\sigma}$ are creation and annihilation operators, respectively, for the state with wave vector $\mathbf{k}$ and spin σ.

Now we introduce the lattice. First of all a static lattice of course scatters single particle states, modulates their wave function, and gives rise to a change in the single particle dispersion relation $H_{\sigma\sigma'}(\mathbf{k})$. In reality it is also the positive ions of the lattice that are responsible for screening out the interaction between the electrons. If we therefore allow for dynamic changes in the lattice, it becomes possible for one electron to deform the lattice through its Coulomb interaction with the positive ions. This deformation in turn leads to a change in interaction between the lattice and the other electrons, which is possible to see as an effective interaction between the electron causing the deformation and the one reacting to it. Lattice deformations away from equilibrium are known as phonons, and if we forget about the underlying lattice, we can think of the effective interaction as a phonon mediated interaction directly between the electrons. Introducing the coupling parameter $V_{\mathbf{k}\sigma\mathbf{k}'\sigma'\mathbf{k}''\sigma''\mathbf{k}'''\sigma'''}$, which describes the induced two-electron scattering amplitudes, the system can be modeled with the Hamiltonian

$$\mathcal{H} = \sum_{\mathbf{k}\sigma\sigma'} H_{\sigma\sigma'}(\mathbf{k}) c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} + \sum V_{\mathbf{k}\sigma\mathbf{k}'\sigma'\mathbf{k}''\sigma''\mathbf{k}'''\sigma'''} c^\dagger_{\mathbf{k}\sigma} c^\dagger_{\mathbf{k}'\sigma'} c_{\mathbf{k}''\sigma''} c_{\mathbf{k}'''\sigma'''}. \tag{4.2}$$

Cooper now considered the case of two electrons of opposite momentum and spin. He showed that no matter how small the interaction $V_{\mathbf{k}\sigma\mathbf{k}'\sigma'\mathbf{k}''\sigma''\mathbf{k}'''\sigma'''}$ is, as long as it is attractive, it will lead to the formation of bound pairs of particles, so called Cooper pairs [62]. Being bound, it takes a finite amount of

energy to break them apart. This effectively prohibits additional perturbative terms added to Eq. (4.2) from inducing single particle scattering, and electrons therefore only scatter in pairs through the potential $V_{\mathbf{k}\sigma\mathbf{k}'\sigma'\mathbf{k}''\sigma''\mathbf{k}'''\sigma'''}$. It is this protection from single particle scattering that is responsible for the resistance free flow in superconductors, as the simultaneous scattering of two electrons in a Cooper pair results in no net change in charge carrier momentum.

Drawing on Coopers observation that electrons can form bound pairs, BCS then imagined a ground state formed from virtual Cooper pairs consisting of electrons of oposite spin and momentum [56]. The only interaction terms that are important in the Hamiltonian are then $V_{\mathbf{k}\uparrow-\mathbf{k}\downarrow-\mathbf{k}'\downarrow\mathbf{k}'\uparrow} \equiv V_{\mathbf{k}\mathbf{k}'}$. They further assumed[1] that these elements can be approximated by $V_{\mathbf{k}\mathbf{k}'} = -V$, where an explicit minus sign has been introduced to indicate that the interaction is attractive [56, 62]. With this background we can now understand why the BCS Hamiltonian that finally proved successful at describing superconductivity is [56]

$$\mathcal{H} = \sum_{\mathbf{k}\sigma\sigma'} H_{\sigma\sigma'}(\mathbf{k}) c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} - V \sum_{\mathbf{k}\mathbf{k}'} c^\dagger_{\mathbf{k}\uparrow} c^\dagger_{-\mathbf{k}\downarrow} c_{-\mathbf{k}'\downarrow} c_{\mathbf{k}'\uparrow}. \tag{4.3}$$

4.1.2 The BCS wave function

Having understood the rational behind the BCS Hamiltonian in Eq. (4.3), we introduce the wave function that describes the ground state. To arrive at this state, we note that we expect the electrons to pair up in virtual Cooper pairs. We therefore write down the most general wave function where all electrons are paired with opposite momentum and spin. This is given by the quantum mechanical superposition [56]

$$|\Psi\rangle = \sum g(\mathbf{k}_{i_1}, ..., \mathbf{k}_{i_n}) c^\dagger_{\mathbf{k}_{i_1}\uparrow} c^\dagger_{-\mathbf{k}_{i_1}\downarrow} ... c^\dagger_{\mathbf{k}_{i_n}\uparrow} c^\dagger_{-\mathbf{k}_{i_n}\downarrow} |0\rangle, \tag{4.4}$$

where $|0\rangle$ is the vacuum and $i_1, ..., i_n$ runs over all combinations of $2n$-particle states. However, the number of coefficients $g(\mathbf{k}_{i_1}, ..., \mathbf{k}_{i_n})$ in this expression is tremendous, and BCS therefore retreated to the use of the wave function [56]

$$|\Psi_{BCS}\rangle = \prod_{\mathbf{k}} \left(u_{\mathbf{k}} + v_{\mathbf{k}} c^\dagger_{\mathbf{k}\uparrow} c^\dagger_{-\mathbf{k}\downarrow} \right) |0\rangle. \tag{4.5}$$

It is clear that $|u_{\mathbf{k}}|^2$ and $|v_{\mathbf{k}}|^2$ are probabilities that the Cooper pair $\mathbf{k}\uparrow -\mathbf{k}\downarrow$ is unoccupied and occupied, respectively. As the pair is either unoccupied or occupied, it is also clear that $|u_{\mathbf{k}}|^2 + |v_{\mathbf{k}}|^2 = 1$.

We make two notes about the assumptions that go into the BCS wave function. First of all, when expanded, it is seen that the number of particles is not

[1] This corresponds to assuming that the interaction is isotropic. The assumption is motivated by the observation that conventional superconductivity is universal over a large class of materials, and therefore should not depend on the specific symmetries of any particular material [56].

well defined, which stands in contrast to the wave function in Eq. (4.4). BCS originally treated the wave function in Eq. (4.5) as a precursor to the physical wave function, which they considered to be the projection of the BCS wave function onto the space of $2n$-particle wave functions [56]. Second, we note that even if we project out the states with n Cooper pairs, the resulting wave function can only describe those states that have coefficients of the form

$$g(\mathbf{k}_{i_1}, ..., \mathbf{k}_{i_n}) = \prod_{\mathbf{k}\in\mathbf{K}} u_{\mathbf{k}} \prod_{\mathbf{k}'\in\mathbf{K}'} v_{\mathbf{k}'}, \tag{4.6}$$

where $\mathbf{K} \cap \mathbf{K}' = \emptyset$, and $\mathbf{K}' = \{\mathbf{k_i}, ..., \mathbf{k}_{i_n}\}$. Although there is no guarantee that this is the case, the justification for this treatment lies in the success of the theory. There is also a second way of interpreting the BCS wave function in Eq. (4.5), which avoids the technicality of projecting out states with fixed particle numbers. In this interpretation it is noted that the BCS wave function is sharply peaked about the correct number of particles, and that it can be seen as corresponding to a grand canonical ensemble of states [59].

4.2 Bogoliubov-de Gennes formalism[2]

4.2.1 Mean field treatment

It is now possible to use variational calculus to solve for the coefficients in the wave function, as BCS originally did [56]. However, an approach that is more suitable for calculations is provided by the Bogoliubov-de Gennes mean field formulation [58, 59, 63, 64]. To arrive at this we recall that the BCS wave function according to one interpretation can be seen as representing a grand canonical ensemble. It is therefore possible to take macroscopic expectation values of the form

$$F_{\mathbf{k}\sigma\sigma'} = \langle c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} \rangle, \tag{4.7}$$

and expect these to be non-zero. The reason why this can be non-zero is that the BCS wave function is a superposition of states with different number of Cooper pairs. The removal of a pair does therefore not kill the expectation value. This corresponds to the statement that superconductivity is a phenomenon whereby a macroscopic number of electrons condenses into a superconducting state filled with a macroscopic number of Cooper pairs. The removal of a single such pair from the condensate should not affect the physics considerably. We will come back with more details on this in Section 4.3.

[2] Part of the discussion in Section 4.2 and 4.3 runs in parallel with the discussion in the section "Particle-hole symmetry" in Paper III.

Let us apply this idea to the BCS Hamiltonian in Eq. (4.3), but assume for generality that the interaction term is of the form

$$\mathcal{H}_{\text{int}} = -\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} V_{\mathbf{k}\mathbf{k}'\sigma\sigma'} c^\dagger_{\mathbf{k}\sigma'} c^\dagger_{-\mathbf{k}\sigma} c_{-\mathbf{k}'\sigma} c_{\mathbf{k}'\sigma'}. \tag{4.8}$$

We can then formally rewrite the interaction term as

$$\mathcal{H}_{\text{int}} = -\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} V_{\mathbf{k}\mathbf{k}'\sigma\sigma'} (F^*_{\mathbf{k}\sigma\sigma'} + c^\dagger_{\mathbf{k}\sigma'} c^\dagger_{-\mathbf{k}\sigma} - F^*_{\mathbf{k}\sigma\sigma'})(F_{\mathbf{k}'\sigma\sigma'} + c_{-\mathbf{k}'\sigma} c_{\mathbf{k}'\sigma'} - F_{\mathbf{k}'\sigma\sigma'}). \tag{4.9}$$

Now consider $F_{\mathbf{k}\sigma\sigma'}$ to be a mean field approximation for $c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'}$, and assume that the variation around the mean value is small. It is then possible to expand the expression above, and discard terms that are quadratic in $c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} - F_{\mathbf{k}\sigma\sigma'}$, to arrive at

$$\mathcal{H}_{\text{int}} = -\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} V_{\mathbf{k}\mathbf{k}'\sigma\sigma'} \left(F_{\mathbf{k}'\sigma\sigma'} c^\dagger_{\mathbf{k}\sigma'} c^\dagger_{-\mathbf{k}\sigma} + F^*_{\mathbf{k}\sigma\sigma'} c_{-\mathbf{k}'\sigma} c_{\mathbf{k}'\sigma'} - F^*_{\mathbf{k}\sigma\sigma'} F_{\mathbf{k}'\sigma\sigma'} \right). \tag{4.10}$$

Noting that $F^*_{\mathbf{k}\sigma\sigma'} F_{\mathbf{k}'\sigma\sigma'}$ only will give a constant shift to the energy, we can discard this term to finally arrive at the mean field Hamiltonian

$$\mathcal{H}_{\text{MF}} = \sum_{\mathbf{k}\sigma\sigma'} H_{\sigma\sigma'}(\mathbf{k}) c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} + \sum_{\mathbf{k}\sigma\sigma'} \left(\Delta_{\mathbf{k}\sigma\sigma'} c^\dagger_{\mathbf{k}\sigma'} c^\dagger_{-\mathbf{k}\sigma} + \Delta^*_{\mathbf{k}\sigma\sigma'} c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} \right). \tag{4.11}$$

Here we have defined the superconducting order parameter

$$\Delta_{\mathbf{k}\sigma\sigma'} \equiv -\sum_{\mathbf{k}'} V_{\mathbf{k}\mathbf{k}'\sigma\sigma'} F_{\mathbf{k}'\sigma\sigma'}. \tag{4.12}$$

The way $F_{\mathbf{k}\sigma\sigma'}$ is defined through Eq. (4.7), we can think of it as the probability that we are able to remove a Cooper pair in the state $|-\mathbf{k}\sigma\mathbf{k}\sigma'\rangle$ from the condensate. That is, the probability that this particular Cooper pair exists. Multiplying by the interaction strength, and summing over all $\mathbf{k}$-states the pair can be in, we see that $\Delta_{\mathbf{k}\sigma\sigma'}$ is related to the energy gained through the formation of such a pair.[3]

[3] Beware that this argument is a little bit oversimplified. Strictly speaking, $F_{\mathbf{k}\sigma\sigma'}$ is a probability that a pair can be removed, while the state still remains in the ground state. It is further seen from Eq. (4.11) that $\Delta_{\mathbf{k}\sigma\sigma'}$ appears as the strength of a source term for Cooper pairs in the state $\mathbf{k}\uparrow -\mathbf{k}\downarrow$. The order parameter $\Delta_{\mathbf{k}\sigma\sigma'}$ is therefore an average over the number of states that scatter into this pair state, ignoring which state these pairs originated from. Such scattering does lower the energy, and is therefore related to the binding energy. This is why we say that $\Delta_{\mathbf{k}\sigma\sigma'}$ is related to the binding energy, rather than being a binding energy. Further, scattering of a Cooper pair can only occur if there are available states to scatter into. Only if we deal

4.2.2 Nambu spinors

Through the introduction of $\Delta_{\mathbf{k}\sigma\sigma'}$, we have succeeded at reducing the Hamiltonian to a quadratic form. We know that quadratic Hamiltonians are comparatively easy to solve, because they can be diagonalized using a single particle basis. However, this is only true when the operators appears in pairs of one creation operator and one annihilation operator. In the Bogoliubov-de Gennes mean field Hamiltonian in Eq. (4.11), creation operators pair up with creation operators, and likewise for annihilation operators. This means that an ordinary single particle treatment is still not possible. It is, however, possible to diagonalize the Hamiltonian using what is known as a Nambu spinor basis

$$(c_{\mathbf{k}\uparrow}, c_{\mathbf{k}\downarrow}, c^\dagger_{-\mathbf{k}\uparrow}, c^\dagger_{-\mathbf{k}\downarrow}). \tag{4.13}$$

Doing so, the problem can be written as [61]

$$\mathcal{H}(\mathbf{k}) = \begin{bmatrix} \mathbf{H}(\mathbf{k}) & \mathbf{\Delta}_{\mathbf{k}} \\ -\mathbf{\Delta}^*_{-\mathbf{k}} & -\mathbf{H}^T(-\mathbf{k}) \end{bmatrix}, \tag{4.14}$$

where $\mathbf{H}(\mathbf{k})$ and $\mathbf{\Delta}_{\mathbf{k}}$ are two-by-two matrices with components $H_{\sigma\sigma'}(\mathbf{k})$ and $\Delta_{\mathbf{k}\sigma\sigma'}$, respectively.[4] The eigenvectors of this matrix will further be written as

$$|\gamma^{(E)}_{\mathbf{k}}\rangle = \begin{bmatrix} u^{(E)}_{\mathbf{k}\uparrow} & u^{(E)}_{\mathbf{k}\downarrow} & v^{(E)}_{\mathbf{k}\uparrow} & v^{(E)}_{\mathbf{k}\downarrow} \end{bmatrix}^T \equiv \begin{bmatrix} \mathbf{u}^{(E)}_{\mathbf{k}} & \mathbf{v}^{(E)}_{\mathbf{k}} \end{bmatrix}^T, \tag{4.15}$$

where E is the eigenvalue of the state.

Although the use of the Nambu spinor basis is helpful for diagonalizing the Hamiltonian, it does come with an additional consequence. The use of a basis consisting of both creation and annihilation operators, means that the apparent degrees of freedom are doubled. More specifically, the number of eigenstates are twice the number of single particle states. However, the number of physical degrees of freedom should of course not change just because of a change of formalism, as should be particularly clear from considering the non-superconducting case $\mathbf{\Delta}_{\mathbf{k}} = 0$. It turns out that although the number of eigenstates is doubled, there is a correspondence between each state at E, to a corresponding one at $-E$ [58]. To see this, consider writing the equation

with fractionally occupied states, as is the case close to the Fermi level for a superconductor, the scattering is allowed to reduce the energy. Consider therefore also two degenerate Cooper pair states with momentum $(\mathbf{k}, -\mathbf{k})$ and $(\mathbf{k}', -\mathbf{k}')$, respectively, which are shared by a single Cooper pair. The interaction term hybridizes these two states and pushes one up in energy and one down in energy, this results in a reduction in the energy for the electron pair that goes into the lower energy state. However, if the two states are both fully occupied there is no energy gain as the gain in the lower level is cancelled by the loss in the higher level.

[4] To connect to Eq. (4.11) write $\sum_{\mathbf{k}} C^\dagger \mathcal{H}(\mathbf{k}) C$, where $C^\dagger = [\ c^\dagger_{\mathbf{k}\uparrow} \quad c^\dagger_{\mathbf{k}\downarrow} \quad c_{-\mathbf{k}\uparrow} \quad c_{-\mathbf{k}\downarrow}\ ]$, perform the matrix multiplication, and permute the operators into appropriate position. The result is Eq. (4.11) up to an additional constant factor that can be ignored, and a relative factor 2 that can be absorbed by rescaling the parameters in the problem.

$\mathcal{H}(\mathbf{k})|\gamma_{\mathbf{k}}^{(E)}\rangle = E|\gamma_{\mathbf{k}}^{(E)}\rangle$ as

$$\mathbf{H}(\mathbf{k})\mathbf{u}_{\mathbf{k}}^{(E)} + \boldsymbol{\Delta}_{\mathbf{k}}\mathbf{v}_{\mathbf{k}}^{(E)} = E\mathbf{u}_{\mathbf{k}}^{(E)}, \tag{4.16}$$

$$-\boldsymbol{\Delta}_{-\mathbf{k}}^{*}\mathbf{u}_{\mathbf{k}}^{(E)} - \mathbf{H}^{T}(-\mathbf{k})\mathbf{v}_{\mathbf{k}}^{(E)} = E\mathbf{v}_{\mathbf{k}}^{(E)}. \tag{4.17}$$

Compare this to the complex conjugate of the expression for $\mathcal{H}(-\mathbf{k})|\gamma_{-\mathbf{k}}^{(-E)}\rangle = -E|\gamma_{-\mathbf{k}}^{(-E)}\rangle$

$$\mathbf{H}^{T}(-\mathbf{k})\mathbf{u}_{-\mathbf{k}}^{(-E)*} + \boldsymbol{\Delta}_{-\mathbf{k}}^{*}\mathbf{v}_{-\mathbf{k}}^{(-E)*} = -E\mathbf{u}_{-\mathbf{k}}^{(-E)*}, \tag{4.18}$$

$$-\boldsymbol{\Delta}_{\mathbf{k}}\mathbf{u}_{-\mathbf{k}}^{(-E)*} - \mathbf{H}(\mathbf{k})\mathbf{v}_{-\mathbf{k}}^{(-E)*} = -E\mathbf{v}_{-\mathbf{k}}^{(-E)*}. \tag{4.19}$$

On inspection we see that these are the same set of equations, meaning that if $\begin{bmatrix} \mathbf{u}_{\mathbf{k}} & \mathbf{v}_{\mathbf{k}} \end{bmatrix}^{T}$ is an eigenstate of $\mathcal{H}(\mathbf{k})$ with eigenvalue E, then it is related to the eigenstate of $\mathcal{H}(-\mathbf{k})$ with energy $-E$ through

$$\begin{bmatrix} \mathbf{u}_{\mathbf{k}}^{(E)} \\ \mathbf{v}_{\mathbf{k}}^{(E)} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{-\mathbf{k}}^{(-E)*} \\ \mathbf{u}_{-\mathbf{k}}^{(-E)*} \end{bmatrix}. \tag{4.20}$$

This is not reflective of an ordinary symmetry that can be broken by a physical perturbation, but rather is a fundamental consequence of using the Nambu-basis. It results in a particle-hole symmetry in the formalism that we will refer to as a BdG-type particle-hole symmetry, to distinguish it from other types of particle-hole symmetries that can be broken.

4.2.3 Reducing the number of degrees of freedom

For the original BCS case where $H_{\sigma\sigma'}(\mathbf{k})$ is diagonal and equal for both spins, say $H(\mathbf{k}) \equiv H_{\uparrow\uparrow}(\mathbf{k}) = H_{\downarrow\downarrow}(\mathbf{k})$, and $\boldsymbol{\Delta}_{\mathbf{k}} = i\Delta\sigma_y$ [61], the Hamiltonian can further be decoupled into two two-by-two matrices of the form

$$\mathcal{H}_{\pm}(\mathbf{k}) = \begin{bmatrix} H(\mathbf{k}) & \pm\Delta \\ \pm\Delta^{*} & -H(\mathbf{k}) \end{bmatrix}, \tag{4.21}$$

now written in the two bases

$$\begin{aligned} &(c_{\mathbf{k}\uparrow}, c_{-\mathbf{k}\downarrow}^{\dagger})_{+}, \\ &(c_{\mathbf{k}\downarrow}, c_{-\mathbf{k}\uparrow}^{\dagger})_{-}. \end{aligned} \tag{4.22}$$

As far as the energy spectrum is concerned, the two problems in Eq. (4.21) are the same,[5] and it is therefore enough to consider the positive sign problem, for

[5] This can be seen by considering the characteristic equation, which in both cases reads $0 = \det(\mathcal{H}_{\pm}(\mathbf{k}) - E) = E^2 - H^2(\mathbf{k}) - |\Delta|^2$.

which we write the eigenvectors as

$$|\Psi\rangle = \begin{bmatrix} u_{\mathbf{k}\uparrow}^{(E)} \\ v_{\mathbf{k}\downarrow}^{(E)} \end{bmatrix}_+ . \tag{4.23}$$

The positive and negative energy eigenstates further have to satisfy

$$\begin{bmatrix} u_{\mathbf{k}\uparrow}^{(E)} \\ v_{\mathbf{k}\downarrow}^{(E)} \end{bmatrix}_+ = \begin{bmatrix} v_{\mathbf{k}\downarrow}^{(-E)*} \\ -u_{\mathbf{k}\uparrow}^{(-E)*} \end{bmatrix}_+ , \tag{4.24}$$

in order to be orthogonal to each other. A similar relation obviously also hold for the negative sign basis where $\uparrow$ and $\downarrow$ are interchanged.

By only considering the positive sign problem in the ordinary BCS problem, the number of eigenstates is halved, and the artificial doubling of degrees of freedom is thereby avoided. This specific choice of reduction of degrees of freedom is particularly useful for finding the quasi-particle energies, as it involves solving a single two-by-two matrix. It is therefore commonly used in treatments of ordinary BCS superconductivity. There is, however, freedom in how to perform the reduction to the correct number of degrees of freedom, and a different choice will turn out to be conceptually useful.

Using Eq. (4.20), it is possible to relate the negative eigenstate solution for the positive sign problem, to the positive eigenvalue solution for the negative sign problem. We can therefore consider only positive solutions, at the expense of having to consider both the positive and negative sign problem. This is appealing, because the eigenstates are closely related to creation and annihilation operators for quasi-particles known as Bogoliubov quasi-particles [58, 61]. Such quasi-particles are excitations on top of the ground state, and therefore have positive energy [58]. In particular, the use of only positive energy eigenstates has the benefit of being generally applicable for any problem of the four-by-four form in Eq. (4.14). We will later also see that the alternative choice of only considering negative eigenstates is useful when considering ground state properties. Additionally, putting the dividing line between positive and negative energy states, rather than between different blocks of the Hamiltonian (which may or may not exist), has the benefit of eventually enabling a very general treatment of Majorana fermions, and we therefore strongly advocate this approach here.

4.2.4 Bogoliubov quasi-particle operators[6]

To make the link to standard notation for the ordinary BCS problem explicit [58, 59], and to eventually be able to describe the connection to the BCS wave

[6] This section is particularly hard to read due to the explicit use of multiple notations for both bases and vector component labels. The reader is therefore advised to read the paragraph with pen and paper, and consult Fig. 4.1.

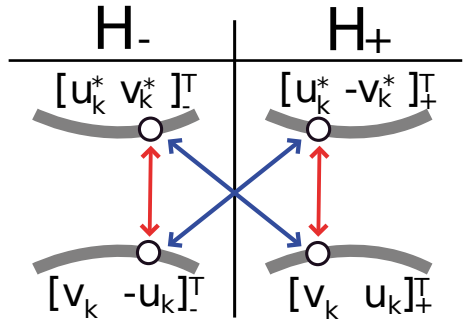

Figure 4.1. In the ordinary BCS problem, which can be treated with a two-by-two matrix, it is common to consider only the positive sign problem in Eq. (4.21), and to consider the states at $\pm E$ to correspond to independent excitations. However, Eq. (4.20) relates positive and negative energy eigenstates along the diagonal blue arrows. That is, it is a relation between positive and negative energy eigenstates for different two-by-two Hamiltonians. This gives us freedom to choose to work with for example only the positive energy eigenstates of the whole problem, rather than to work with the positive and negative energy eigenstates of the positive sign problem only. We here advocate the former approach. We also note that the coefficients of the eigenstates at positive and negative energy for $\mathcal{H}_+$ or $\mathcal{H}_-$ are related along the red vertical arrows through Eq. (4.24). This is however not a relation between equivalent states, but rather between orthogonal eigenvectors.

function and its excitations, let the negative energy eigenstate for $\mathcal{H}_+(\mathbf{k})$ be denoted by $[v_{\mathbf{k}} \;\; u_{\mathbf{k}}]^T_+$.[7] In the full four-by-four basis in Eq. (4.13), this state can also be written as $[v_{\mathbf{k}} \;\; 0 \;\; 0 \;\; u_{\mathbf{k}}]^T$. Using Eq. (4.24), the positive energy eigenstate for $\mathcal{H}_+(\mathbf{k})$ is then seen to be given by $[u^*_{\mathbf{k}} \;\; -v^*_{\mathbf{k}}]^T_+$. Similarly, using the four-by-four basis expression together with Eq. (4.20), the positive energy eigenstate for $\mathcal{H}_-(-\mathbf{k})$ is seen to be given by $[0 \;\; u^*_{\mathbf{k}} \;\; v^*_{\mathbf{k}} \;\; 0]^T$, which in the two-by-two basis is $[u^*_{\mathbf{k}} \;\; v^*_{\mathbf{k}}]^T_-$. The positive energy eigenstates for $H_+(\mathbf{k})$ and $H_-(-\mathbf{k})$ can therefore in this notation be written as

$$
\begin{aligned}
|\Psi_+(\mathbf{k})\rangle &= \left[\begin{array}{cc} u^*_{\mathbf{k}} & -v^*_{\mathbf{k}} \end{array} \right]^T_+ , \\
|\Psi_-(-\mathbf{k})\rangle &= \left[\begin{array}{cc} u^*_{\mathbf{k}} & v^*_{\mathbf{k}} \end{array} \right]^T_- .
\end{aligned}
\tag{4.25}
$$

Through the negative sign analogue of Eq. (4.24), the negative energy eigenstate for $\mathcal{H}_-(-\mathbf{k})$ is in the same way given by $[v_{\mathbf{k}} \;\; -u_{\mathbf{k}}]^T_-$. See Fig. 4.1 for a depiction of the relation between the coefficients of the different eigenstates. We emphasize that the coefficients in this notation, and the coefficients in the

[7] The labels for the components may seem arbitrary but have been chosen with hindsight. The justification for this particular choice has to do with the particular form the Bogoliubov quasiparticle states acquire, and their relation to the coefficients in the BCS wave function. In particular, the usage of the same symbols as in Eq. (4.5) will be justified in Section 4.3.2.

notation given by Eq. (4.23), are related through[8]

$$\begin{aligned} u_{\mathbf{k}} &= v_{\mathbf{k}\downarrow}^{(-E)} = u_{\mathbf{k}\uparrow}^{(E)*} = -v_{-\mathbf{k}\uparrow}^{(-E)} = u_{-\mathbf{k}\downarrow}^{(E)*}, \\ v_{\mathbf{k}} &= u_{\mathbf{k}\uparrow}^{(-E)} = -v_{\mathbf{k}\downarrow}^{(E)*} = u_{-\mathbf{k}\downarrow}^{(-E)} = v_{-\mathbf{k}\uparrow}^{(E)*}. \end{aligned} \tag{4.26}$$

Explicitly writing the positive energy eigenstates in Eq. (4.25) using the operator basis in Eq. (4.22), these define operators[9]

$$\begin{aligned} \gamma_{\mathbf{k}\uparrow}^{\dagger} &\equiv u_{\mathbf{k}}^{*} c_{\mathbf{k}\uparrow}^{\dagger} - v_{\mathbf{k}}^{*} c_{-\mathbf{k}\downarrow}, \\ \gamma_{-\mathbf{k}\downarrow}^{\dagger} &\equiv u_{\mathbf{k}}^{*} c_{-\mathbf{k}\downarrow}^{\dagger} + v_{\mathbf{k}}^{*} c_{\mathbf{k}\uparrow}. \end{aligned} \tag{4.27}$$

These are creation operators for the Bogoliubov quasi-particles mentioned above [58, 59]. We note that the choice of $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ as symbols for the components of the eigenstates is suggestive of them being related to the components of the BCS wave function in Eq. (4.5), for which the same symbols have been used. This is indeed the case, and we will in the next section demonstrate why this is true for the ordinary BCS case. We will also see in the next section how the Bogoliubov quasi-particle operators create excitations on top of the ground state, but for the moment we simply assume that this is the case. Further, we note that the negative energy eigenstates $[v_{\mathbf{k}} \;\; u_{\mathbf{k}}]_{+}$ and $[v_{\mathbf{k}} \;\; -u_{\mathbf{k}}]_{-}$, similarly as above define operators

$$\begin{aligned} \tilde{\gamma}_{\mathbf{k}\uparrow}^{\dagger} &\equiv v_{\mathbf{k}} c_{\mathbf{k}\uparrow}^{\dagger} + u_{\mathbf{k}} c_{-\mathbf{k}\downarrow} = \gamma_{-\mathbf{k}\downarrow}, \\ \tilde{\gamma}_{-\mathbf{k}\downarrow}^{\dagger} &\equiv v_{\mathbf{k}} c_{\mathbf{k}\downarrow}^{\dagger} - u_{\mathbf{k}} c_{-\mathbf{k}\uparrow} = \gamma_{\mathbf{k}\uparrow}. \end{aligned} \tag{4.28}$$

In the rightmost equalities, we used that the middle expressions are seen to be Hermitian conjugates of the expressions in Eq. (4.27). It is therefore clear that it is formally possible to consider the creation of a quasi-particle above $E = 0$ to be the same as the annihilation of the corresponding state below $E = 0$, and vice versa. This is another manifestation of the doubling of the degrees of freedom and the BdG-type particle-hole symmetry mentioned above: both the creation and annihilation operators are explicitly encoded in the eigenstates. For an ordinary Hamiltonian, only the creation operator is explicitly encoded

[8] These relations follow from writing down the four eigenstates using both notations, and identifying coefficients. With the notation from Eq. (4.24) used on the right hand side of the equality sign, the four eigenstates read

$$\begin{aligned} [v_{\mathbf{k}} \;\; u_{\mathbf{k}}]_{+}^{T} &= [u_{\mathbf{k}\uparrow}^{(-E)} \;\; v_{\mathbf{k}\downarrow}^{(-E)}]_{+}^{T} \qquad [u_{\mathbf{k}}^{*} \;\; -v_{\mathbf{k}}^{*}]_{+}^{T} = [u_{\mathbf{k}\uparrow}^{(E)} \;\; v_{\mathbf{k}\downarrow}^{(E)}]_{+}^{T} \\ [v_{\mathbf{k}} \;\; -u_{\mathbf{k}}]_{-}^{T} &= [u_{-\mathbf{k}\downarrow}^{(-E)} \;\; v_{-\mathbf{k}\uparrow}^{(-E)}]_{-}^{T} \qquad [u_{\mathbf{k}}^{*} \;\; v_{\mathbf{k}}^{*}]_{-}^{T} = [u_{-\mathbf{k}\downarrow}^{(E)} \;\; v_{-\mathbf{k}\uparrow}^{(E)}]_{-}^{T} \end{aligned}$$

[9] Note that this is written using the conjugate transpose of the basis in Eq. (4.22), because the eigenvectors are column vectors.

in the eigenstates, while the annihilation operator is obtained as the Hermitian conjugate of the eigenstates.[10]

Generally the positive energy eigenstates for the four-by-four Hamiltonian defines quasi-particle operators

$$\gamma_{\mathbf{k}}^{(E)\dagger} = u_{\mathbf{k}\uparrow}^{(E)} c_{\mathbf{k}\uparrow}^{\dagger} + u_{\mathbf{k}\downarrow}^{(E)} c_{\mathbf{k}\downarrow}^{\dagger} + v_{\mathbf{k}\uparrow}^{(E)} c_{-\mathbf{k}\uparrow} + v_{\mathbf{k}\downarrow}^{(E)} c_{-\mathbf{k}\downarrow}, \tag{4.29}$$

which through Eq. (4.20) are related to quasi-particle operators below $E = 0$ according to

$$\gamma_{\mathbf{k}}^{(E)\dagger} = \gamma_{-\mathbf{k}}^{(-E)}. \tag{4.30}$$

This relation will turn out to be particularly useful when we turn to a discussion about Majorana fermions in Chapter 5.

4.3 The particle-hole picture

4.3.1 Addition and removal of Cooper pairs

Let us now pause for a moment to interpret what the Bogoliubov-de Gennes formulation, and eigenstates associated with the Nambu spinors mean. To construct the BCS wave function, we assumed that it is reasonable to treat the problem by only having an approximate knowledge of the number of Cooper pairs in the system. The rationale is that superconductivity arise as a macroscopic number of Cooper pairs form. The removal of a single pair should therefore not make the macroscopic state qualitatively different [33]. This allows us to assume that it is possible to define an order parameter $\Delta_{\mathbf{k}\sigma\sigma'}$, which characterizes the ground state, and is related to the number of Cooper pairs. The reason why the uncertainty in the number of particles in the BCS wave function is important for $\Delta_{\mathbf{k}\sigma\sigma'}$ to exist, is due to the fact that the expectation value in Eq. (4.7) is identically zero for a ground state with a fixed particle number. However, in reality the number of particles is of course conserved as long as we ignore interactions with the environment. We also strongly suspect that interactions with the environment should not be a prerequisite for superconductivity. At the same time, if $\Delta_{\mathbf{k}\sigma\sigma'}$ is related to the number of Cooper pairs, then it clearly ought to exist in a state that is formed from a fixed number of Cooper pairs. It therefore seems as if we need to find a way to reconcile these conflicting ideas.

[10] We note that mathematically it is possible to say that the creation operators for the eigenstates of an ordinary single-particle Hamiltonian corresponds to the right eigenvectors of the Hamiltonian, while the annihilation operators correspond to the left eigenvectors. In particular, the creation and annihilation operators can be considered basis vectors for the column and row vectors, respectively, according to the identification $c_{\mathbf{k}\uparrow}^{\dagger} = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$, $c_{\mathbf{k}\downarrow} = \begin{bmatrix} 0 & 1 \end{bmatrix}$, and so forth. However, in Nambu space both $c_{\mathbf{k}\sigma}^{\dagger}$ and $c_{\mathbf{k}\sigma}$ corresponds to basis vectors for both the row and column vectors.

One way to interpret the BCS wave function is to think of the fixed particle number state $|\phi + pair\rangle$ with a Cooper pair added to the fixed particle number state $|\phi\rangle$ as the same state as $|\phi\rangle$. As argued above, as far as the other particles in the condensate are concerned, their state should not be affected to any relevant extent by such addition or removal of a pair. This argument means that we in principle could calculate $\Delta_{\mathbf{k}\sigma\sigma'}$ also using a physical (particle number conserving) wave function such as Eq. (4.4). To do so would only require a modification of the process whereby a scalar product between states such as $|\phi + pair\rangle$ and $|\phi\rangle$ is taken, in such a way as to accordingly give finite results, as long as the two states only differ in number of Cooper pairs. As we will see below, the BCS wave function can be seen as a technical construct that allows for exactly this.

4.3.2 Quasi-particle excitations and the BCS ground state

We now demonstrate how the technical details of the discussion above work out in practice for the ordinary BCS case, described by Eq. (4.21). In addition, this will help us understand the claim in Section 4.2.4, that the Bogoliubov quasi-particles are excitations from the ground state. Consider therefore the Bogoliubov quasi-particle operators $\gamma_{\mathbf{k}\uparrow}$ and $\gamma_{-\mathbf{k}\downarrow}$ introduced in Eq. (4.27). When acting on the ground state, these operators both create and destroy particles. Formally they appear as a mixture of particles and holes. Assume now that the coefficients in the BCS wave function in Eq. (4.5) indeed are the same as the coefficients in $\gamma_{\mathbf{k}\uparrow}$, as hinted at in Section 4.2.2. We will soon see why this assumption is correct. Applying the $\gamma_{\mathbf{k}\uparrow}$-operator to the BCS wave function we obtain

$$\gamma_{\mathbf{k}\uparrow}|\Psi_{BCS}\rangle = 0, \tag{4.31}$$

$$\gamma^\dagger_{\mathbf{k}\uparrow}|\Psi_{BCS}\rangle = c^\dagger_{\mathbf{k}\uparrow}|\Psi^{\mathbf{k}}_{BCS}\rangle, \tag{4.32}$$

$$\gamma_{\mathbf{k}\uparrow}\gamma^\dagger_{\mathbf{k}\uparrow}|\Psi_{BCS}\rangle = |\Psi_{BCS}\rangle. \tag{4.33}$$

Here

$$|\Psi^{\mathbf{k}}_{BCS}\rangle = \prod_{\mathbf{k}'\neq\mathbf{k}} \left(u_{\mathbf{k}'} + v_{\mathbf{k}'} c^\dagger_{\mathbf{k}'\uparrow} c^\dagger_{-\mathbf{k}'\downarrow}\right)|0\rangle, \tag{4.34}$$

is the BCS wave function with the Cooper pair $\mathbf{k}\uparrow -\mathbf{k}\downarrow$ removed. Corresponding relations can be obtained for $\gamma_{-\mathbf{k}\downarrow}$ by replacing $\mathbf{k}\uparrow$ by $-\mathbf{k}\downarrow$ in all expressions above.

What did just happen? Apparently $\gamma_{\mathbf{k}\uparrow}$ kills the ground state, while in contrast $\gamma^\dagger_{\mathbf{k}\uparrow}$ removes one Cooper pair by leaving a single unpaired electron behind. Even more intriguingly $\gamma_{\mathbf{k}\uparrow}$ is able to recreate the original ground state after a Cooper pair is replaced by a stray particle. Beginning with the state $\gamma^\dagger_{\mathbf{k}\uparrow}|\Psi_{BCS}\rangle$

we see that, compared to the ground state, this is an excited state. This is clear, because if the ground state prefers to have a certain number of Cooper pairs, then removing one of these pairs and replacing it with a single particle clearly is an excitation. Further, $\gamma_{\mathbf{k}\uparrow}$ is annihilating that same excitation. However, if the excitation is not there it kills the ground state. This is analogous to how ordinary creation and annihilation operators act on the vacuum. We therefore see that the states that arise from the Bogoliubov-de Gennes Hamiltonian describe the relevant quasi-particles that occur on top of the underlying superconducting "vacuum". This also demonstrates that the coefficients in the BCS wave function indeed are related to the coefficients of the excitations as assumed, because by this choice the ground state becomes the state that has no excitations.

Although this clearly shows that the Bogoliubov quasi-particles created by $\gamma^{\dagger}_{\mathbf{k}\uparrow}$ are no stranger than ordinary particles, the calculation above at first seems a little odd. How is it that when we apply an operator that simultaneously removes and adds particles, we end up with a state as simple as that with a single stray particle on top of a condensate of Cooper pairs? To understand this we examine the situation a bit closer. If we before we apply the quasi-particle operators to the state imagine that we were to expand it, we know that we would end up with a superposition of states with different particle numbers. When the quasi-particle operator $\gamma^{\dagger}_{\mathbf{k}\uparrow}$ in Eq. (4.27) now act on this state, it results in the removal of a single electron in the state $-\mathbf{k}\downarrow$ from those basis states that have a Cooper pair in the state $\mathbf{k}\uparrow -\mathbf{k}\downarrow$. However, to those basis states that do not have any Cooper pair in that particular state, an electron is instead simply added in the state $\mathbf{k}\uparrow$. This means that there is a transfer of states both up and down in particles numbers, but all these states afterwards have in common that they have a single unpaired particle in the state $\mathbf{k}\uparrow$. The BCS wave function can therefore simply be seen as a technical construct. It allows the exact number of Cooper pairs to be irrelevant in calculations, at the same time as it gives the same physical picture as a fixed particle number wave function. In reality we know that the actual physical ground state is of the form in Eq. (4.4), and that it can be projected out of the BCS wave function [56]. But, because we are not interested in the ground state wave function itself as much as we are interested in quantities like $\Delta_{\mathbf{k}\sigma\sigma'}$ and the excitations, the BCS wave function can for all practical purposes itself be considered to be the ground state. Similarly, we note that although the Bogoliubov quasi-particle operators formally appear as mixtures of electrons and holes, it is a bit misleading to think of the excitations they create as mixtures of electrons and holes. The excitations they create are clearly unpaired electrons, not a superposition of electrons and holes. The operators should therefore rather be understood to, in conspiracy with the BCS wave function, simplify the mathematics.

4.3.3 The explicit BCS wave function is superfluous

We are now ready to demonstrate that the explicit construction of the BCS wave function is not required. However, before arriving at this conclusion, we start by assuming that we have been given the correct ground state in the form of the BCS wave function. We then proceed to calculate the pair function $F_{\mathbf{k}\uparrow\downarrow}$ using Eq. (4.7), which results in

$$F_{\mathbf{k}\downarrow\uparrow} = \langle \Psi_{BCS} | c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow} | \Psi_{BCS} \rangle = \langle \Psi^{\mathbf{k}}_{BCS} | u^*_{\mathbf{k}} v_{\mathbf{k}} | \Psi^{\mathbf{k}}_{BCS} \rangle = u^*_{\mathbf{k}} v_{\mathbf{k}}. \quad (4.35)$$

Here we have limited ourselves to $F_{\mathbf{k}\downarrow\uparrow}$, because this is the only pair-function that is relevant for ordinary BCS theory, which was considered above. The equation tells us that there is a very simple relation between the pair function and the coefficients of the Bogoliubov-de Gennes eigenstates. This shows that the knowledge of the Nambu spinor coefficients is enough to allow us to side step the BCS wave function when calculating quantities such as the pair function and order parameter $\Delta_{\mathbf{k}\sigma\sigma'}$. As we will see in the next section, this has important implications for the general case where the matrix is four-by-four, and for which no explicit wave function is provided.

4.3.4 The Bogoliubov-de Gennes band structure

As mentioned in Section 4.2.3, and demonstrated in Section 4.3.2, it is beneficial to consider the positive eigenstates as encoding creation operators for excitations. In Section 4.3.2 we also saw that the corresponding annihilation operators, which are related to the creation operators encoded in the negative energy eigenstates through Eq. (4.30), kills the ground state. The negative eigenstates can therefore be considered as corresponding to occupied quasiparticle states in the ground state. This is very similar to how ordinary ground states consist of band structures with all states below the Fermi level occupied, and all above unoccupied.[11]

There are two important differences between a normal band structure and the Bogoliubov-de Gennes band structure. First, the Fermi level will always sit at $E = 0$. Of course the Fermi level of the BCS Hamiltonian in Eq. (4.3) can vary, but the chemical potential enters in the dispersion relation $H_{\sigma\sigma'}(\mathbf{k})$. This means that the effect of a shift in Fermi energy modifies the band structure, which follows from the Bogoliubov-de Gennes Hamiltonian in Eq. (4.21), rather than shift the line that divides occupied and unoccupied states from each

[11] This conclusions may seem trivial. However, the truth is that the Bogoliubov-transformation [57–59] as it is usually done obscures some of these simple facts. For example, it is common to calculate ground state properties for a superconductor by summing over both positive and negative eigenstates with a strangely looking occupation factor of $2n(E)-1$, rather than just the Fermi-Dirac distribution $n(E)$ as is the case for an ordinary band structure. This discussion has been carried out for $T = 0$, but for finite temperatures the result is to receive the familiar occupation factor $n(E)$ rather than $2n(E) - 1$.

other. We can say that the chemical potential sets the Fermi level for the normal state described by $H_{\sigma\sigma'}(\mathbf{k})$, while E in the quasi-particle Hamiltonian refers to energies relative to the Fermi level. Second, we note that in an ordinary band structure particle and hole excitations can occur independently of each other. However, we know from Eq. (4.30), that in the Bogoliubov-de Gennes band structure the deoccupation of a level below $E = 0$ necessarily comes together with the occupation of a corresponding state above $E = 0$.

To take advantage of this physically intuitive picture, we use Eq. (4.26) to express the pair function in Eq. (4.35) as

$$F_{\mathbf{k}\downarrow\uparrow} = \langle c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow} \rangle = v_{\mathbf{k}\downarrow}^{(-E)*} u_{\mathbf{k}\uparrow}^{(-E)}. \tag{4.36}$$

Further, $|u_{\mathbf{k}}|^2$ and $|v_{\mathbf{k}}|^2$ are defined as the probabilities that the pair $\mathbf{k}\uparrow -\mathbf{k}\downarrow$ is unoccupied and occupied, respectively. This means that it also is the probability that the state $\mathbf{k}\uparrow$ is unoccupied and occupied, respectively. It is therefore clear that the coefficients $|u_{\mathbf{k}\uparrow}^{(-E)}|^2 = |v_{\mathbf{k}}|^2$ and $|v_{\mathbf{k}\downarrow}^{(-E)}|^2 = |u_{\mathbf{k}}|^2$ have the same meaning. In particular, the expression for the pair function above aquires the form of an ordinary expectation value, where $c_{\mathbf{k}\uparrow}$ acts as an annihilation operator to the right, extracting the probability amplitude for that state being occupied. Similarly, $c_{-\mathbf{k}\downarrow}$ acts as a creation operator to the left, extracting the probability amplitude for that the state is unoccupied. This observation allows us to generalize the expression for the pair function to

$$F_{\mathbf{k}\sigma\sigma'} = \langle c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} \rangle = \sum_{E_\lambda < 0} v_{\mathbf{k}\sigma}^{(\lambda)*} u_{\mathbf{k}\sigma'}^{(\lambda)}. \tag{4.37}$$

Here the sum is over all states below $E = 0$, to take into account that the probability to find the particle is the sum over the probability to find it in any given band λ. We emphasise that Eq. (4.37) allows us to calculate the pair function in the absence of an explicit BCS like wave function, which we do not have in the general case.

4.4 Superconducting mass gap and superflow

By now we know how to interpret the Bogoliubov-de Gennes Hamiltonian and its eigenstates, but what are the energies of the excitations? To answer this we have to diagonalize the Bogoliubov-de Gennes Hamiltonian. We will later have to do so for a general four-by-four matrix, but here we restrict ourselves to the ordinary two-by-two BCS Hamiltonian in order to demonstrate the concept of a superconducting mass gap. A comparison between Eq. (4.21) and Eq. (3.1) reveals that we have already solved this problem, with the result

$$E = \pm\sqrt{|\Delta|^2 + H^2(\mathbf{k})}. \tag{4.38}$$

It is clear that the dispersion relation always has an energy gap around $E = 0$. Further, because all states below $E = 0$ are occupied, while all above are empty, the superconductor is in many ways similar to a band insulator. However, the superconductor is not an insulator, quite the opposite.

To understand in what sense superconductors are similar to insulators, and in what sense they differ, we perform a Galilean boost of the electrons. If this were to happen in a vacuum, Galilean invariance would guarantee this motion to continue indefinitely. That is, we would have a constant current that does not die off. However, this argument also holds for electrons that are not in a superconducting state. The difference between the two fluids becomes clear when we admit that in reality the electrons in a conductor are moving within a lattice. The lattice is still in the lab frame, and this breaks the Galilean invariance, because although kinetic energy of unidirectional motion is relative, friction forces make the lab frame the preferred coordinate system in which to minimize the free energy. The non-moving electron fluid therefore becomes the energetically favorable condition.

A normal fluid eventually falls back into the low energy state, because scattering against impurities continuously redistributes the momentum. This is where a superconductor is different. Because the electrons are bound to each other inside the superconducting condensate, scattering of single particles would need to break up Cooper pairs. This costs energy and therefore the individual particles become insensitive to scattering. Impurities can attempt to scatter particles, but the condensate ensures that the particles remain in the superflow.[12] The only way to change the flow in the superconductor is therefore to act on the condensate as a whole with a macroscopic force [56]. This is what happens when an external electromagnetic field acts on a macroscopic piece of the superconductor. We see that the superconductor is like an insulator in the sense that it costs a large amount of energy to create an excitation out of the ground state. The superconductor is, however, different from the insulator in that the superconducting ground state itself can be superflowing.

[12] Considering that the attractive electron-electron interaction is mediated by phonons, it appears to me that the appropriate explanation for what actually happens is the following. A particle, say $\mathbf{k}\uparrow$, can in fact scatter against an impurity into a state $\mathbf{q}\uparrow$. However, this impurity is fixed to the lattice, and sets up a phonon of momentum $\mathbf{k}-\mathbf{q}$ in order to conserve momentum. The phonon is then absorbed by the electron in the state $-\mathbf{k}\downarrow$, which accordingly is scattered into the state $-\mathbf{q}\downarrow$. The result is that scattering do occur, but only through process that do not leak momentum (and energy) into the environment, and therefore preserves the momentum of the condensate as a whole. To ensure that momentum and energy is not leaked into the environment, the phonons should not be able to propagate freely in the material. This is exactly the case for the phonons that mediates the electron-electron interaction, as they are virtual phonons [65]. That is, they are modes that decays exponentially because they do not carry enough energy to support free propagation.

4.5 Self-consistent method

We have seen in Section 4.3.2 that $\gamma_{\mathbf{k}}^{(\lambda)\dagger} = \gamma_{-\mathbf{k}}^{(-\lambda)}$, for $E_\lambda > 0$, is an operator that creates excitations from the ground state. We also learned in Section 4.3.4 that the coefficients $u_{\mathbf{k}\sigma}^{(\lambda)}$ and $v_{\mathbf{k}\sigma}^{(\lambda)}$ for $E_\lambda < 0$ determine the ground state. In particular, the coefficients can be used to calculate the pair function $F_{\mathbf{k}\sigma\sigma'}$ and order parameter $\Delta_{\mathbf{k}\sigma\sigma'}$, through Eqs. (4.37) and (4.12), respectively. The order parameter in turn determines the Bogoliubov-de Gennes Hamiltonian, from which the eigenstates $\gamma_{\mathbf{k}}^{(\lambda)\dagger}$ follows. This suggests that it is possible to self-consistently determine the ground state. That is, for any given $H_{\sigma\sigma'}(\mathbf{k})$ and pair potential $V_{\mathbf{k}\mathbf{k}'\sigma\sigma'}$, it should be possible to determine its superconducting ground state if we can find a Δ_{Trial} such that

$$\Delta_{\text{Trial}} \stackrel{?}{=} \Delta_{\text{BdG}} \equiv -\sum_{\mathbf{k}'} V_{\mathbf{k}\mathbf{k}'\sigma\sigma'} F_{\mathbf{k}'\sigma\sigma'} = -\sum_{\mathbf{k}', E_\lambda < 0} V_{\mathbf{k}\mathbf{k}'\sigma\sigma'} v_{\mathbf{k}'\sigma}^{(\lambda)*} u_{\mathbf{k}'\sigma'}^{(\lambda)}. \quad (4.39)$$

Here Δ_{Trial} is the order parameter which determines the Hamiltonian, while Δ_{BdG} follows from the eigenstates of the Hamiltonian. The $\mathbf{k}\sigma\sigma'$-indices are here implicit for the Δ-symbols.

The equation above describes the self-consistency condition, but also gives a recipe for how to arrive at a self-consistent solution. By repeating the sequence

$$\Delta_{\text{Trial}}^n \to \mathcal{H}(\mathbf{k}) \to \Delta_{\text{BdG}}^n \equiv \Delta_{\text{Trial}}^{n+1}, \quad (4.40)$$

using some reasonable initial Δ_{Trial}^0, and in each step perform the self-consistency test

$$||\Delta_{\text{Trial}}^n - \Delta_{\text{BdG}}^n|| < \epsilon, \quad (4.41)$$

we can for a given error tolerance ϵ eventually arrive at a self-consistent result. Here $||\Delta||$ denotes some appropriate norm on the space of order parameters, such as for example $||\Delta|| = \max_{\mathbf{k}\sigma\sigma'} |\Delta_{\mathbf{k}\sigma\sigma'}|$. Although the procedure usually is very stable, it is important to be aware that such a self-consistent procedure can lead to many different solutions, corresponding to different local minima in the free energy. In fact, we will in what follows be interested in solutions corresponding to both global and local minima in the form of homogenous bulk and vortex solutions, respectively.

4.6 Real space formulation

4.6.1 Transformed expressions

We have so far worked in momentum space, but it will be particularly important for us to be able to work in real space. Further, we will mainly need to consider the real space expression for the ordinary BCS case in Eq. (4.3).

More specifically, we will consider the case where $V_{\mathbf{k}\mathbf{k}'} = -V$, and for which it suffices to consider the single matrix elements $F_{\mathbf{k}} = F_{\mathbf{k}\downarrow\uparrow} = -F_{\mathbf{k}\uparrow\downarrow}$ and $\Delta = \Delta_{\mathbf{k}\downarrow\uparrow} = -\Delta_{\mathbf{k}\uparrow\downarrow}$ of the pair function and order parameter, respectively. To arrive at the real space formulation, we expand the operators in the previous expressions in their real space basis, and perform the sums over momentum. The result for the single particle term is dependent on the form of $H_{\sigma\sigma'}(\mathbf{k})$, and we therefore leave this for later. However, the interaction term in the Bogoliubov-de Gennes Hamiltonian, the order parameter, and the pair function, in Eq. (4.7, 4.11, 4.12) becomes

$$\begin{aligned}\mathcal{H}_{\text{int}} =& \sum_{\mathbf{x}} \left(\Delta c^\dagger_{\mathbf{x}\uparrow} c^\dagger_{\mathbf{x}\downarrow} + \Delta^* c_{\mathbf{x}\downarrow} c_{\mathbf{x}\uparrow} \right), \\ \Delta =& -V \sum_{\mathbf{x}} F_{\mathbf{x}}, \\ F_{\mathbf{x}} =& \langle c_{\mathbf{x}\downarrow} c_{\mathbf{x}\uparrow} \rangle. \end{aligned} \tag{4.42}$$

Once we have these expressions, the same doubling of degrees of freedom and introduction of Nambu spinors can be done as in momentum space. In real space the single particle part of the Bogoliubov-de Gennes Hamiltonian will however not be diagonal. This means that instead of a set of decoupled four-by-four matrices, we will need to write down a matrix of dimension $2 \times 2 \times n$, where n is the number of sites on the lattice under consideration. The first factor of two comes from having two spins at each site, while the second factor of two is due to the artificial doubling of the number of degrees of freedom.

We also note that while the present derivation was performed in the homogeneous case, and Δ comes out as a sum over all space, we will in real space calculations promote Δ to a space dependent variable $\Delta_{\mathbf{x}}$. This is done by assuming the order parameter to be local in the sense that $\Delta_{\mathbf{x}} = -(V/N)F_{\mathbf{x}} \to -VF_{\mathbf{x}}$, where the number of lattice sites N is absorbed by the pair potential. The expression for the order parameter above will thereby be interpreted as saying that the average of the order parameter over all sites is the same as the order parameter at any particular site when the system is homogeneous. We can alternatively say that this corresponds to the assumption that the pair potnetial is completely local in space.

4.6.2 Superflow in real space

Consider now a Galilean boost $\mathbf{k} \to \mathbf{k} + \frac{\mathbf{q}}{2}$ on all the eigenstates of the Bogoliubov-de Gennes Hamiltonian. Further, assume that the boost is small enough that $|u^{(\lambda)}_{\mathbf{k}+\frac{\mathbf{q}}{2}\sigma}| \approx |u^{(\lambda)}_{\mathbf{k}\sigma}|$. Then the only effect is that the components of the eigenstates that go into the calculation in Eq. (4.37) are multiplied by a factor[13] $e^{i\mathbf{q}\cdot\mathbf{x}/2}$.

[13] Note that the the $v_{\mathbf{k}\sigma}$ components are multiplied by a factor $e^{-i\mathbf{q}\cdot\mathbf{x}/2}$, because these are associated with the operator $c^\dagger_{-\mathbf{k}\sigma}$, rather than $c_{\mathbf{k}\sigma}$. In particular, the boost should be understood to

This results in the pair function being multiplied by an overall factor of $e^{i\mathbf{q}\cdot\mathbf{x}}$, and this carries over into the real space expressions for the pair function in Eq. (4.42) and to the local order parameter $\Delta_{\mathbf{x}}$, as there is no summation over $\mathbf{q}$. This means that when the ground state experiences a superflow with $\frac{\mathbf{q}}{2}$ flow per particle, the order parameter becomes position dependent and is given by

$$\Delta_{\mathbf{x}} = \Delta e^{i\mathbf{q}\cdot\mathbf{x}}. \tag{4.43}$$

We therefore see that a supercurrent is associated with a change of phase in the order parameter. It is also clear that the momentum per Cooper pair at $\mathbf{x}$ can be obtained from the order parameter through the expression

$$q^{\mu} = \mathrm{Im}\left(\frac{\Delta^{*}_{\mathbf{x}}\partial_{\mu}\Delta_{\mathbf{x}}}{|\Delta_{\mathbf{x}}|^{2}}\right). \tag{4.44}$$

4.6.3 Vortices and associated equivalence classes

The expression in Eq. (4.44) for the Cooper pair momentum looks very much like the connections we saw in the previous two chapters. This suggests that there may be topological properties associated with it, and there are indeed. However, instead of applying the connection above in order to classify different equivalence classes, we here use more physical arguments to understand what the objects of these topological equivalence classes are.

From a physical point of view, we can in general expect that the superflow instead of being homogeneous as above, varies from point to point in space. This means that it is possible to imagine a superflow that gives rise to an order parameter of the form

$$\Delta_{\mathbf{x}} = \Delta(\rho, z)e^{in\theta}, \tag{4.45}$$

where (ρ, θ, z) are the ordinary cylindrical coordinates, and n is some integer. This is clearly a superflow that is circulating around $\mathbf{x} = 0$, and corresponds to a vortex in a superconductor [66]. Using Eq. (4.44) we also see that the superflow is getting quicker and quicker the closer to $\rho = 0$ we are. As we approach the center of the vortex, the superflow becomes infinite, which clearly is not a physical possibility. However, this type of flow is possible if we at the same time assume that the superconductivity vanishes at the center, such that $\Delta(\rho, z) \to 0$ as $\rho \to 0$. In a sense a vortex can therefore be seen as a possible configuration of the order parameter in a superconductor that has a hole drilled in it.

We now note that the requirement that n is an integer is important. If it were to be any other real number, the order parameter would have to make a discontinuous jump somewhere around the vortex. This would make the derivative as

mean $-\mathbf{k} \to -\mathbf{k} + \frac{\mathbf{q}}{2}$ for the lower block of the Hamiltonian, because after a boost electrons with momentum $\mathbf{k} + \frac{\mathbf{q}}{2}$ are paired with electrons with momentum $-\mathbf{k} + \frac{\mathbf{q}}{2}$.

well as superflow infinite across this jump, which is unphysical. It is therefore clear that it is impossible to continuously transform a vortex with some n, into a vortex with a different n. It is, however, possible to continuously deform a vortex with some n, to some other configuration of the order parameter which also winds n times around $\rho = 0$. The set of vortices having the form given above can therefore from a topological point of view be seen as a representation of all possible single vortex configurations. In this sense a vortex is an example of a topological excitation [3, 5].

Because $\Delta_{\mathbf{x}}$ is a complex number defined over real space, we can think of the problem as a fiber bundle with a three-dimensional base manifold and a complex fiber. However, a closer look at Eq. (4.44) reveals that it is ill-defined at $\rho = 0$. A simpler picture therefore emerge if we instead take the view expressed above, that a vortex is a configuration of the order parameter in a superconductor with a hole drilled in it. Each node in the order parameter is responsible for drilling another hole. The topological equivalence classes[14] that arise from a given number of holes are then represented by the vortex configurations that winds $n_1, ..., n_i$ times around hole 1 to i. This is an example of how not only the object (the connection or order parameter), but also the embedding space (the fiber bundle), is responsible for what the possible equivalence classes are. This is similar to the case of the threaded donut in Fig. 2.6.

4.7 Unconventional superconductivity

In this chapter we have introduced a general formalism for treating superconductors. However, we have several times limited our discussion to special cases. For example, when discussing superflow in Section 4.4, the discussion was limited to the ordinary BCS problem in the form of a two-by-two matrix. In this case it is implied that only $\Delta_{\mathbf{k}\downarrow\uparrow} = -\Delta_{\mathbf{k}\uparrow\downarrow}$ and $F_{\mathbf{k}\downarrow\uparrow} = -F_{\mathbf{k}\uparrow\downarrow}$ are non-zero. Similarly, when discussing the real space formulation in Section 4.6.1, which was performed for a full four-by-four matrix, the discussion was still limited to the same conditions on the order parameter and pair function. The superconductors considered in these special cases are known as conventional s-wave superconductors. To understand why, and to introduce the notion of unconventional superconductivity, such as p-wave superconductivity, we take a closer look at the pair function.

Using the fermionic commutation relations, we see that the pair function satisfies the relation

$$F_{\mathbf{k}\sigma\sigma'} = \langle c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'} \rangle = -\langle c_{\mathbf{k}\sigma'} c_{-\mathbf{k}\sigma} \rangle = -F_{-\mathbf{k}\sigma'\sigma}. \tag{4.46}$$

[14] This is assuming that the vortex lines themselves do not form knots. In that case one can imagine many more configurations.

This tells us that the pair function is odd under the simultaneous reversal of momentum and interchange of spin indices. We may, however, also ask how the pair function transforms under the individual actions of either reversing the momentum or interchanging the spin indices. For example, in the cases where we assumed $F_{\mathbf{k}\uparrow\downarrow} = -F_{\mathbf{k}\downarrow\uparrow}$, it is clear that the pair function is odd under interchange of spin. We now write the pair function for a given $\mathbf{k}$ on the matrix form

$$\mathbf{F}_{\mathbf{k}} \equiv \begin{bmatrix} F_{\mathbf{k}\uparrow\uparrow} & F_{\mathbf{k}\uparrow\downarrow} \\ F_{\mathbf{k}\downarrow\uparrow} & F_{\mathbf{k}\downarrow\downarrow} \end{bmatrix} \equiv F_{\mathbf{k}0}\sigma_0 + F_{\mathbf{k}x}\sigma_x + F_{\mathbf{k}y}\sigma_y + F_{\mathbf{k}z}\sigma_z, \tag{4.47}$$

where σ_i are the three Pauli matrices and the identity. In the last step we used that any complex two-by-two matrix can be written as a sum of the Pauli matrices and the identity, with complex coefficients $F_{\mathbf{k}i}$. It is clear that these coefficients are given by

$$\begin{aligned} F_{\mathbf{k}0} =& \frac{1}{2}\langle c_{-\mathbf{k}\uparrow}c_{\mathbf{k}\uparrow} + c_{-\mathbf{k}\downarrow}c_{\mathbf{k}\downarrow}\rangle, \\ F_{\mathbf{k}x} =& \frac{1}{2}\langle c_{-\mathbf{k}\uparrow}c_{\mathbf{k}\downarrow} + c_{-\mathbf{k}\downarrow}c_{\mathbf{k}\uparrow}\rangle, \\ F_{\mathbf{k}y} =& \frac{i}{2}\langle c_{-\mathbf{k}\uparrow}c_{\mathbf{k}\downarrow} - c_{-\mathbf{k}\downarrow}c_{\mathbf{k}\uparrow}\rangle, \\ F_{\mathbf{k}z} =& \frac{1}{2}\langle c_{-\mathbf{k}\uparrow}c_{\mathbf{k}\uparrow} - c_{-\mathbf{k}\downarrow}c_{\mathbf{k}\downarrow}\rangle. \end{aligned} \tag{4.48}$$

Here $F_{\mathbf{k}0}, F_{\mathbf{k}x}$ and $F_{\mathbf{k}z}$ are said to be coefficients that describe triplet Cooper pairing, because they are derived from operators that act on the triplet spin space, which is spanned by $|\uparrow\uparrow\rangle$, $|\downarrow\downarrow\rangle$ and $(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)/\sqrt{2}$. In contrast, $F_{\mathbf{k}y}$ describes spin singlet pairing, because it is derived from an operator that acts on the singlet spin space spanned by $(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$. It is therefore possible to classify superconductors into spin singlet and spin triplet superconductors, depending on which coefficients that are non-zero [61].

To further investigate the properties of the singlet and triplet superconductors, we note that the interchange of spin indices corresponds to taking the transpose of $\mathbf{F}_{\mathbf{k}}$, which results in

$$\mathbf{F}_{\mathbf{k}}^T = F_{\mathbf{k}0}\sigma_0 + F_{\mathbf{k}x}\sigma_x - F_{\mathbf{k}y}\sigma_y + F_{\mathbf{k}z}\sigma_z. \tag{4.49}$$

The spin singlet coefficient is therefore odd under the interchange of spin, while the triplet coefficients are even. However, because Eq. (4.46) can be written as

$$\mathbf{F}_{\mathbf{k}} = -\mathbf{F}_{-\mathbf{k}}^T, \tag{4.50}$$

it is clear that all coefficients are odd when both $\mathbf{k}$ is reversed, and spins are interchanged. In order for the total oddness to be preserved, this leads to the

conclusion that as a function of $\mathbf{k}$, the singlet pair function must be even, while the triplet pair functions must be odd.

Expanding the $F_{\mathbf{k}i}$ coefficients in the basis of spherical harmonics, they can further be written as

$$F_{\mathbf{k}i} = \sum_{nlm} c^i_{nlm} j_n(|\mathbf{k}|) Y_{lm}(\theta, \varphi). \tag{4.51}$$

Here $j_n(|\mathbf{k}|)$ are the radially dependent parts of the expansion, while $Y_{lm}(\theta, \varphi)$ are the spherical harmonics. The parameters $|\mathbf{k}|$, θ and φ are the spherical coordinates of $\mathbf{k}$. Now spherical harmonics are even or odd under parity depending on whether l is even or odd [67]. It is therefore clear that the single pair function $F_{\mathbf{k}y}$ only contains terms with even l. Following the terminology from atomic physics, the singlet pair function is said to be $s, d, g, ...$-wave. Similarly, the triplet pair functions $F_{\mathbf{k}0}$, $F_{\mathbf{k}x}$, and $F_{\mathbf{k}z}$, only contains terms with odd l, and are therefore labelled $p, f, h, ...$-wave.

The same properties are inherited by the order parameter, which is calculated from the pair function. However, it is common to express the order parameter using a slightly different notation [61]

$$\mathbf{\Delta}_{\mathbf{k}} = i \left(\psi(\mathbf{k}) + \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\sigma} \right) \sigma_y, \tag{4.52}$$

where $i\psi(\mathbf{k})\sigma_y$ is the singlet component, $\mathbf{d}(\mathbf{k})$ is a complex vector, and $\boldsymbol{\sigma}$ is a vector of Pauli matrices. Having the same properties as the pair function, it is clear that also $\psi(\mathbf{k})$ and $\mathbf{d}(\mathbf{k})$ can be expanded as

$$\psi(\mathbf{k}) = \sum_{nlm} c^{\psi}_{nlm} j_n(|\mathbf{k}|) Y_{lm}(\theta, \varphi), \tag{4.53}$$

$$d_i(\mathbf{k}) = \sum_{nlm} c^{i}_{nlm} j_n(|\mathbf{k}|) Y_{lm}(\theta, \varphi), \tag{4.54}$$

where $\psi(\mathbf{k})$ consists of terms with even l only, and the components of $\mathbf{d}(\mathbf{k})$ consists of terms with odd l only.

4.8 Superconductivity in two dimensions

In this thesis we are mainly concerned with superconductivity in two dimensions. While reality of course always have three spatial dimensions, two-dimensional effective models can arise in many ways. For the systems we consider there are two main routes for motivating such a model. First, such a model can be motivated when studying the surface of a three-dimensional bulk superconductor. In this case the effective two-dimensional model arise by projecting out the part of the Hamiltonian that describes the motion along the surface. The perpendicular component is assumed to be possible to integrate

out, and that this at most give rise to a renormalization of the model parameters for the remaining Hamiltonian. Second, it is also possible to consider two dimensional structures that are not superconducting in themselves, but which are in contact with a superconducting substrate. A superconductor has a certain superconducting coherence length, which is the length scale over which a superconductor transitions from superconducting to non-superconducting [59]. When a material is placed in contact with a superconductor, it is therefore possible that superconductivity is not destroyed at the interface, but instead leaks into the nearby material [68]. As far as the model we consider is concerned, both of these motivations are valid. Our results are therefore independent of which particular experimental realization that is most promising.

Finally, we mention that when the system is two-dimensional the coordinate θ in Eq. (4.53) becomes fixed at $\theta = \frac{\pi}{2}$, and only the variable φ becomes relevant. For any fixed θ, the spherical harmonics satisfy

$$Y_{lm}(\theta, \varphi) \sim e^{im\varphi}, \tag{4.55}$$

and with a slight abuse of notation we correspondingly consider states with $m = 0, 1, 2, ...$ to be $s, p, d, ...$-wave superconductors, respectively.

5. Topological superconductivity

In the previous chapter we introduced the Bogoliubov-de Gennes formalism for superconductors, and saw that this leads to a band picture of superconductivity. We also saw that this band structure is formally similar to a band insulator. This suggests that the topological band theory in Chapter 3 is applicable also to superconductors. Moreover, we saw in the previous chapter that a vortex can be seen as a topological excitation in real space. In this chapter we will see how the interplay between these real and momentum space topologies can give rise to what is known as localized Majorana fermions in vortex cores. We will also see that such Majorana fermions can appear at endpoints of one-dimensional wires. However, we will also see that the term Majorana fermion is slightly misleading. Majorana non-Abelion may in fact may be a more appropriate name.

5.1 Majorana fermion

In Section 3.2 we noted the structural similarities between the dispersion relation associated with a small or vanishing mass gap, and the relativistic dispersion relation that follows from the Dirac equation. This led to a formal analogy between the electron and hole excitations in condensed matter, and the electron and positron excitations that are solutions to the Dirac equation. In 1937, the Italian physicist Ettore Majorana proposed an alternative equation for relativistic particles, for which particles are their own antiparticles [32, 33, 69]. That is, the process that creates a particle is the same as that which annihilates it, $\gamma^\dagger = \gamma$. This is reminiscent of the relations between the Bogoliubov quasi-particles, where Eq. (4.30) reads

$$\gamma_{\mathbf{k}}^{(E)\dagger} = \gamma_{-\mathbf{k}}^{(-E)}. \tag{5.1}$$

The analogy is however not complete, because although this means that the creation of one state is the same as the annihilation of another, it is not the same state that is created and annihilated. The relation above only states that the occupation of a state above $E = 0$ is accompanied by a deoccupation of a state below $E = 0$.

Before moving on with the analogy, we note here that Eq. (5.1) is derived using the four-by-four matrix formulation of the Bogoliubov-de Gennes Hamiltonian in Section 4.2.2. We stress that the relation does not hold between the positive and negative eigenstates in the ordinary BCS Hamiltonian,

when, as is customary, this problem is formulated using only the positive sign problem in Eq. (4.21). In this case Eq. (5.1) relates eigenstates in the two problems with different sign to each other, as indicated in Fig. 4.1. This possible source of confusion is one reason for why it has been argued that it is beneficial to always treat the full four-by-four problem, and to consider only positive eigenstates as corresponding to excitations. We will therefore here always start from the complete four-by-four formulation.

Although Eq. (5.1) does not in itself imply a Majorana relation, it is possible to get one step closer by imagining a Dirac cone in the energy spectrum. In this case $E = 0$ at $\mathbf{k} = 0$, and the relation $\gamma_0^{(0)\dagger} = \gamma_0^{(0)}$ follows.[1] Moreover, because of the bulk-boundary correspondence presented in Sections 3.3 and 3.4, single or multiple branches of Dirac cones can be expected to occur on edges of two-dimensional superconductors. It is only required that the superconductor is described by a Hamiltonian that gives rise to a non-trivial Chern number. However, because the Bogoliubov-de Gennes Hamiltonian in general is a four-by-four matrix, it is possible to arrive at a similar situation to that of the topological insulator described in Section 3.4. That is, the two operators are on two different helical branches, and that there is still a hidden band index that separates the two operators. This is not too surprising, as it is possible that the two superscripts in the relation $\gamma_0^{(0)\dagger} = \gamma_0^{(0)}$ only appear to be the same because $E = -E$.

With this background we understand that a non-degenerate zero energy state is a sure sign of a Majorana fermion, because there are then no possible hidden indices. However, at first this seems hopeless to achieve, because we know from Section 4.2.2 that Bogoliubov quasi-particles necessarily come in pairs. The relation between positive and negative energy states is a consequence of the artificial doubling of degrees of freedom, which guarantees each state to have a corresponding partner. In particular, any state with energy $E = 0$ therefore has to have a partner with energy $E = 0$. In spite of this it turns out that it is possible to separate these pairs in space, and it is therefore possible to have locally non-degenerate zero energy states. We will see how below, but before continuing with this analysis we pause to say a few words about the current status of the field.

[1] In general the Dirac point does not need to be located at $\mathbf{k} = 0$, but can be at any high symmetry point in the Brillouin zone. However, as long as there is a single Dirac cone in the system, it needs to be at a high symmetry point where $\mathbf{k}$ is identified with $-\mathbf{k}$ through a reciprocal lattice vector. We therefore use $\mathbf{k} = 0$ as a short hand notation for indicating a high symmetry point, for which $\mathbf{k}$ and $-\mathbf{k}$ refers to the same wave vector. Alternatively, $\mathbf{k}$ can be considered to be the momentum relative to the high symmetry point.

5.2 Theoretical proposals and experimental progress

Several ways to achieve locally non-degenerate zero energy modes have been proposed. Early proposals include vortices in superfluid Helium and spinless superconductors, fractional vortices in fractional quantum Hall states and p-wave superconductors, and at the end points of one-dimensional spinless p-wave superconductors [34, 70, 71]. Not all of these are superconductors, but a common feature of most of them is the use of some type of superfluid state. Particularly, we note the occurrence of spinless and p-wave superconductors in this list, which we will describe in more detail below. None of the above proposals have, however, had any experimental success so far.

Given the difficulty to physically realize any of the above proposals, it is encouraging that a new wave of proposals have followed the discovery of topological insulators. Beginning with the proposal by Fu and Kane to realize effectively spinless superconductors on the surface of a topological insulator [23], a string of other proposals followed suit. Most notably, the first possible signatures of Majorana fermions were reported in 2012, in InSb wires placed on top of an s-wave superconductor [31]. This particular setup is based on several proposals of using the proximity effect to induce s-wave superconductivity into semiconductors with strong Rashba spin-orbit interaction, and further apply a magnetic field to make them effectively spinless [25–28]. The merit of this approach lies in the types of materials that are used to build them. With s-wave superconductors and semiconductors both being materials with a long history of experimental and even industrial progress, this approach has opened up a credible way of experimentally accessing the Majorana physics. More recent experiments have also indicated that ferromagnetic impurities on top of an s-wave superconductor also can provide a promising route toward engineering Majorana fermions.[35–37] In this case the surface is treated as an effectively two-dimensional superconductor, with Rashba spin-orbit interaction arising from inversion symmetry breaking at the surface, while the magnetic impurities provide a Zeeman term. Because both systems involve superconductivity, Rashba spin-orbit interaction, and magnetism, they can both be described by the same model Hamiltonian. From the theoretical point of view presented in this thesis, there is therefore little need for distinguishing between the two systems. The model Hamiltonian that forms the basis for the result part of this thesis can therefore equally well be thought to describe one or the other experimental realization.

5.3 Spinless p-wave superconductor

5.3.1 p-wave superconductor

Although we will be interested in a different setup eventually, it is instructive to consider the spinless p-wave superconductor mentioned above. By doing so

we can understand what is meant by spinless, why it leads to the possible existence of Majorana fermions, as well as prepare for a treatment of the system we eventually will consider. In particular, we consider p-wave superconductivity in a two-dimensional system. We therefore remind ourselves from Section 4.8, that in a two-dimensional system, we mean by p-wave superconductivity that the order parameter varies as

$$\Delta_{\mathbf{k}} \sim e^{-i\varphi}, \tag{5.2}$$

where φ is the polar angle of the wave vector $\mathbf{k} = (|\mathbf{k}|\cos(\varphi), |\mathbf{k}|\sin(\varphi))$. There is still freedom in the radial behavior of the order parameter, but for concreteness we choose an order parameter that varies as

$$\Delta_{\mathbf{k}} \sim k_x - ik_y. \tag{5.3}$$

Further, from Section 4.7 we know that p-wave pairing can only happen in the triplet pairing state. The order parameter therefore has to be described by one, or several, of the components of $\mathbf{d}(\mathbf{k})$ in Eq. (4.52). We choose $\mathbf{d}(\mathbf{k}) = (0, \Delta(k_x - ik_y), 0)$, such that

$$\Delta_{\mathbf{k}} = \begin{bmatrix} \Delta(k_y + ik_x) & 0 \\ 0 & \Delta(k_y + ik_y) \end{bmatrix}, \tag{5.4}$$

where the parameter Δ has been introduced as a parameter through which the strength of the superconductivity can be tuned. To write down the Bogoliubov-de Gennes Hamiltonian, we also need to specify the normal state dispersion relation $H_{\sigma\sigma'}(\mathbf{k})$. For this we choose

$$H(\mathbf{k}) = \begin{bmatrix} \mathbf{k}^2 - \mu + V_z & 0 \\ 0 & \mathbf{k}^2 - \mu - V_z \end{bmatrix}, \tag{5.5}$$

which describes two ordinary parabolic dispersion relations, with a tunable chemical potential μ, and a Zeeman field V_z that can split the two bands. The Bogoliubov-de Gennes Hamiltonian in Eq. (4.14) is then given by

$$\mathcal{H} = \begin{bmatrix} \mathbf{k}^2 - \mu + V_z & 0 & \Delta(k_y + ik_x) & 0 \\ 0 & \mathbf{k}^2 - \mu - V_z & 0 & \Delta(k_y + ik_x) \\ \Delta(k_y - ik_x) & 0 & -\mathbf{k}^2 + \mu - V_z & 0 \\ 0 & \Delta(k_y - ik_x) & 0 & -\mathbf{k}^2 + \mu + V_z \end{bmatrix}. \tag{5.6}$$

Using the two bases

$$(c_{\mathbf{k}\uparrow}, c^\dagger_{-\mathbf{k}\uparrow}), \tag{5.7}$$

$$(c_{\mathbf{k}\downarrow}, c^\dagger_{-\mathbf{k}\downarrow}), \tag{5.8}$$

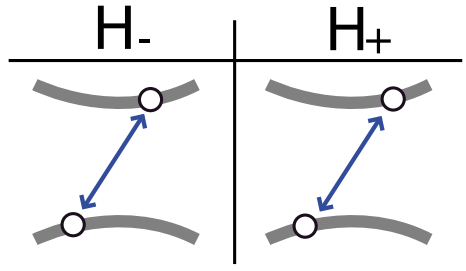

Figure 5.1. The p-wave superconductor described by Eq. (5.6) can be divided into two subblocks $\mathcal{H}_+$ and $\mathcal{H}_-$, described by Eq. (5.9). Further, Eq. (4.20) relates eigenstates along the blue arrows. The relations are between states described by the same submatrix, and any non-degenerate zero energy mode for any of the matrices is therefore related to itself. This stands in contrast to the situation for the ordinary BCS Hamiltonian depicted in Fig. 4.1, where the relation is between states in different submatrices, and therefore is guaranteed to not be related to itself through Eq. (5.1).

the total Hamiltonian can further be split into two independent Hamiltonians of the form

$$\mathcal{H}_\pm(\mathbf{k}) = \begin{bmatrix} \mathbf{k}^2 - \mu \pm V_z & \Delta(k_y + ik_x) \\ \Delta(k_y - ik_x) & -\mathbf{k}^2 + \mu \mp V_z \end{bmatrix}. \tag{5.9}$$

Here the plus and minus sign refers to the problem using the up and down spin basis, respectively. This is similar to how the ordinary BCS problem splits in two different problems, but there is one very important difference. Equation (4.20) relates components associated with up spin, to other components associated with up spin, and likewise for down spins. Any singly degenerate zero energy state for any of the two sub matrices will therefore satisfy the Majorana condition $\gamma^\dagger = \gamma$, because in this case the relation is between states described by the same submatrix. See Fig. 5.1.

5.3.2 Spinless p-wave superconductor

Before continuing with an analysis of possible zero energy states, we reduce the Hamiltonian to an effectively spinless form. To do so we let the chemical potential and Zeeman term be tuned together such that $V_z + \mu \ll -|\Delta|$, while $v_z = \mu - V_z$ remains a small parameter. The down spin problem with negative sign in Eq. (5.9) is then dominated by the diagonal term, and there are no states close to $E = 0$ at all. The down spins therefore have become irrelevant, and it is possible to view the problem as an effectively spinless problem. Of course the electrons have not lost their spins, but there is only one spin species left to consider that is relevant: the up spins. Spinless should therefore not be

misinterpreted to mean that the particles have no spin, but rather that there only is a single relevant spin index.[2]

5.3.3 Edge states

The Hamiltonian that describes the remaining up spin in Eq. (5.9) is now given by

$$\mathcal{H}_+(\mathbf{k}) = \begin{bmatrix} v_z + k_x^2 + k_y^2 & \Delta(k_y + ik_x) \\ \Delta(k_y - ik_x) & -v_z - k_x^2 - k_y^2 \end{bmatrix}. \tag{5.10}$$

A glance at Eq. (3.4) reveals that, with the identification $v_z = M$ and $\Delta = \alpha$, the two Hamiltonians are the same. Formally the p-wave order parameter plays the role of a Rashba term, while v_z is a mass gap. It was there shown that this Hamiltonian has a non-zero Chern number associated with it when $v_z < 0$. Further, the edge states of an infinite strip were considered there. It was shown that along each edge, the Hamiltonian gives rise to a single branch in the dispersion relation which cuts through $E = 0$. Because a single branch cuts through zero, it follows that the state at $E = 0$ is a Majorana fermion. Its partner is localized at the other edge.

5.3.4 Edges with finite length[3]

It seems as if Majorana fermions are bound to appear at the edge of a two-dimensional spinless p-wave superconductor. There is a complication though, because in reality there are no infinite edges, only finite edges that form closed loops. If the infinite strip in Fig. 3.4 is cut so that it has edges on each four sides, all edges are joined into one single edge. Now the fact that there is one branch on each edge before the edges are joined into one, does not imply that there are two branches on the joined edge. From Section 3.4 we remember that the branches on each side of the infinite strip are counter propagating, they therefore propagate in the same direction around the interior. That is, the two branches join into a single clockwise or counter-clockwise propagating branch when the edges are joined. Still, we expect each state to have a partner, but how is this possible if globally there is just a single branch that cuts through zero on a single edge? The answer is that although the branch cuts through zero, the edge now has a finite length, and its spectrum is therefore quantized. This means that the branch is not cutting through $E = 0$ continuously, and there is in fact no state at $E = 0$ [72]. Instead, there are two states just above and below zero that are each others partner. See Fig. 5.2 for a conceptual picture.

[2] More generally the spin index can also refer to a pseudo-spin.

[3] Part of the discussion in this section runs in parallel with the discussion in the section "Majorana spin-polarization and block edge currents" in Paper V.

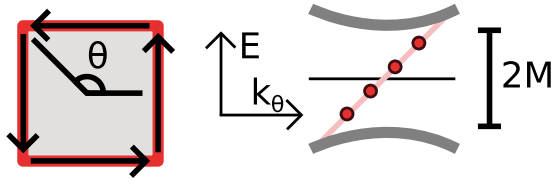

Figure 5.2. When the infinite strip in Fig. 3.4 is cut off, a single edge is formed. The two branches on the opposite sides of the infinite strip join into a single branch on the single edge. Further, the finite length of the edge leads to a discrete energy spectrum. The result is that the branch that cuts through zero does not have a mode with exactly $E = 0$, and there is therefore no Majorana fermion in the system.

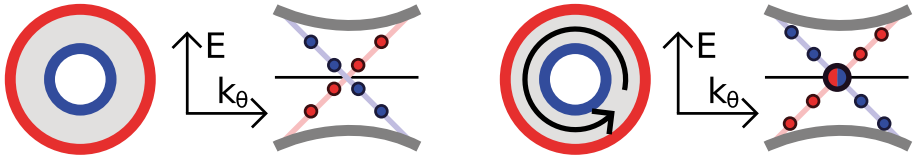

Figure 5.3. (Left) Similar to the situation in Fig. 5.2, but with two edges. Each edge hosts a discretized edge mode branch, but there is no Majorana fermion at $E = 0$. (Right) A magnetic flux is threaded through the center of the disk, or a supercurrent is circulating around it. The discretized spectrum is shifted, such that two Majorana fermions occur at $E = 0$. The two Majorana fermions are located at the inner and the outer edge, respectively.

In spite of the additional problem that arise because of finitely sized edges, it turns out that this problem can be cured. This can be done by drilling a hole into the superconductor, and threading a magnetic flux through the hole. In this way one $E = 0$ state is localized on the inner and outer edge respectively. We do not go into details about this here, but instead refer the reader to [72]. We note, however, that this is equivalent to inserting a vortex into the system. A single vortex in a two-dimensional superconductor therefore gives rise to a Majorana fermion localized at the vortex core, and its partner is localized at the edge of the sample. See Fig. 5.3 for a conceptual picture.

Before moving on we also note that the model described here can be turned into a one-dimensional problem, by ignoring the dimension along the infinite strip, say k_x. The model then reduces to the continuum version of Kitaev's one-dimensional spinless p-wave superconducting wire [34]. This is of particular interest, because Kitaev's one-dimensional spinless p-wave superconductor is an archetypal model, to which the proposals for one-dimensional wires in Section 5.2 are closely related. Kitaev's one-dimensional wire can therefore be seen as the infinite strip with momentum restricted to $k_x = 0$ along the strip. It is therefore clear that although the Chern number cannot be used to

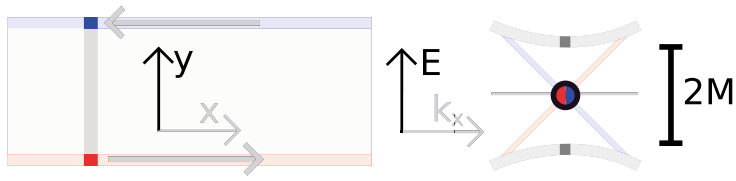

Figure 5.4. By making the restriction $k_x = 0$, the two-dimensional spinless p-wave superconducting infinite strip, can be thought of as a one-dimensional wire along the y-direction. In particular, there are two Majorana fermions at opposite edges of the infinite strip, at $k_x = 0$, which becomes Majorana fermions at opposite ends of the one-dimensional wire. Through this construction, Kitaev's spinless p-wave superconducting wire can be arrived at from a two-dimensional system, which can be classified using a Chern number.

classify the one-dimensional bulk of Kitaev's model,[4] the occurrence of Majorana fermions at the wire's end points is related to a Chern number. That is, the Chern number of the two-dimensional problem from which the one-dimensional model is derived from through dimensional reduction. See Fig. 5.4 for a conceptual picture.

5.4 s-wave superconductor, with Rashba spin-orbit interaction and ferromagnetism

5.4.1 Motivation

The p-wave superconductor described in the previous section captures all the essential details of the type of topological superconductor that we are focused on in the results part. From a theoretical point of view it is therefore an ideal system, because concepts can be easily demonstrated. However, it suffer from the serious drawback that it is hard to realize experimentally. For this reason the model that we consider in the results part, as mentioned in Section 5.2, instead involves conventional s-wave superconducivity [24, 25, 29, 30].

We saw in the previous section that p-wave superconductivity is responsible for providing a Rashba like term in the Hamiltonian in Eq. (5.9). When moving to an s-wave superconductor, the Rashba-like behavior therefore has to come from some other term, and in the setup described here the Rashba interaction itself is used. Apart from s-wave superconductivity and a Rashba term there is, however, no difference in what goes into the construction of the Hamiltonian, as compared to the p-wave superconductor above. We will therefore need s-wave superconductivity, Rashba spin-orbit interaction, and

[4] The first Chern number is trivially zero for the one-dimensional case due to the fact that a one-dimensional manifold cannot be curved. That is, there are to few spatial degrees of freedom for any curvature, such as the Berry curvature in Eq. (2.29), to have a non-zero component.

a Zeeman term in the Hamiltonian, in addition to the ordinary kinetic energy and chemical potential. None of these properties are exotic on their own. It is therefore reasonable to expect that given the readily available amount of materials that exhibits one or several of these effects, it is possible to engineer effectively two-dimensional structures in which all of the effects coexist. As was mentioned in Section 5.2, experimental progress has already been made in this direction. The one-dimensional wire mentioned there is the dimensionally reduced version of the two-dimensional setup. In the same way as Kitaev's model is the dimensionally reduced version of the two-dimensional spinless p-wave superconductor.

5.4.2 Tight-binding model Hamiltonian

Having introduced the system, it is time to specify the model. First of all we will work in a two-dimensional tight-binding model on a square lattice. All three phenomena will be assumed to exist within this layer. Because we work in a tight-binding model, kinetic terms are given by $k_x \to \sin(k_x)$ and $k_x^2 \to 1-\cos(k_x)$. Using this together with our knowledge of the form of the Rashba spin-orbit interaction from Eq. (3.3), we can conclude that the single particle dispersion relation is [30]

$$H_0(\mathbf{k}) = \epsilon(\mathbf{k}) - V_z\sigma_z - \mathcal{L}_0(\mathbf{k})\cdot\boldsymbol{\sigma}, \tag{5.11}$$

$$\epsilon(\mathbf{k}) = -2t\left(\cos(k_x)+\cos(k_y)\right) - \mu, \tag{5.12}$$

$$\mathcal{L}_0(\mathbf{k}) = \alpha(\sin(k_y), -\sin(k_x), 0). \tag{5.13}$$

Here V_z is the strength of the Zeeman term, t is a nearest neighbor hopping parameter, μ the chemical potential, and α the strength of the Rashba spin-orbit interaction. Working in relative units we set $t = 1$ and $\mu \approx \pm 4t$. Including superconductivity and applying the Nambu spinor formalism from Section 4.2.2, we end up with a Hamiltonian in k-space of the form

$$\mathcal{H} = \begin{bmatrix} H_0(\mathbf{k}) & i\Delta\sigma_y \\ -i\Delta^*\sigma_y & -H_0^T(-\mathbf{k}) \end{bmatrix}. \tag{5.14}$$

From Section 4.6.1 we know that the interaction term in an s-wave superconductor in real space takes the form

$$\mathcal{H}_{\text{int}} = \sum_{\mathbf{x}} \left(\Delta_{\mathbf{x}} c^\dagger_{\mathbf{x}\uparrow} c^\dagger_{\mathbf{x}\downarrow} + \Delta^*_{\mathbf{x}} c_{\mathbf{x}\downarrow} c_{\mathbf{x}\uparrow}\right), \tag{5.15}$$

$$\Delta_{\mathbf{x}} = -V\langle c_{\mathbf{x}\downarrow} c_{\mathbf{x}\uparrow}\rangle, \tag{5.16}$$

where V is the strength of the superconducting pair potential. In particular, we have included the possibility for $\Delta_{\mathbf{x}}$ to vary in space. As we have seen in Section 4.6.3 this allows us to study vortices, and it also allows for the self-consistent solution to relax to different values at different points in the lattice,

which is important around edges and impurities. In fact, we also allow for the Zeeman term to be both position dependent and to point in other directions than along the z-axis in some of our works. This allows for ferromagnetic structures such as point impurities or one-dimensional wires embedded in a larger two-dimensional superconductor to be studied. However, here we restrict ourselves to the simpler case of a uniform Zeeman term along the z-direction. Finally, Fourier transforming also the single particle part of the Hamiltonian we arrive at

$$\mathcal{H}_0 = \mathcal{H}_{\text{kin}} + \mathcal{H}_{V_z} + \mathcal{H}_{\text{SO}}, \tag{5.17}$$

$$\mathcal{H}_{\text{kin}} = -t \sum_{\langle \mathbf{i},\mathbf{j}\rangle\sigma} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{j}\sigma} - \mu \sum_{\mathbf{i}\sigma} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{i}\sigma}, \tag{5.18}$$

$$\mathcal{H}_{V_z} = -V_z \sum_{\mathbf{i}\sigma\sigma'} (\sigma_z)_{\sigma\sigma'} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{i}\sigma'}, \tag{5.19}$$

$$\mathcal{H}_{\text{SO}} = -\frac{\alpha}{2} \sum_{\mathbf{ib}} \left(e^{i\theta_{\mathbf{b}}} c^\dagger_{\mathbf{i}+\mathbf{b}\downarrow} c_{\mathbf{i}\uparrow} + \text{H.C.} \right). \tag{5.20}$$

Here $\mathbf{b}$ is a vector along the nearest neighbor bonds, and $\theta_{\mathbf{b}}$ its polar coordinate.

5.4.3 Band structure

Drawing from our experience with two-by-two Hamiltonians in Chapter 3, we begin with an investigation of the bulk band structure. By diagonalizing the Hamiltonian in Eq. (5.14) we arrive at the dispersion relation

$$E(\mathbf{k}) = \pm \sqrt{A(\mathbf{k}) \pm 2\sqrt{B(\mathbf{k})}}, \tag{5.21}$$

$$A(\mathbf{k}) = \epsilon^2(\mathbf{k}) + |\mathcal{L}(\mathbf{k})|^2 + |\Delta|^2 + V_z^2, \tag{5.22}$$

$$B(\mathbf{k}) = \epsilon^2(\mathbf{k})|\mathcal{L}(\mathbf{k})|^2 + (|\Delta|^2 + \epsilon^2(\mathbf{k}))V_z^2. \tag{5.23}$$

This is clearly a more complicated band structure than the two overlapping parabolas studied in Chapter 3. Nevertheless, it is possible to identify some similarities in their structure. We see that both $|\Delta|^2$ and V_z^2 are reminiscent of the mass gap parameters encountered earlier. Although, it is not entirely clear to what extent they play this role in this case. Moreover, the appearance of $\pm 2\sqrt{B(\mathbf{k})}$ seems to complicate things further. A closer look at $\pm 2\sqrt{B(\mathbf{k})}$ reveals that if V_z^2 becomes sufficiently large, and the plus sign is chosen inside the square root, then the Zeeman term dominates the dispersion relation. In the limit $V_z^2 \to \infty$, two of the bands described by Eq. (5.21) therefore are split of from the low energy spectrum around $E = 0$. This leads to an effectively spinless system. However, choosing the negative sign, it is not clear what the increase or decrease of any of the parameters does to the band structure around $E = 0$.

Figure 5.5. Band structure for the model described by Eq. (5.14), for the parameters $t = 1, \mu = 4, \alpha = 2, \Delta = 1$. From left to right $V_z = \frac{\Delta}{2}, \Delta, \frac{3\Delta}{2}$. Note that for $V_z = \Delta$, a Dirac cone appears. On the right hand side the red and green bands are sufficiently split off to make the band structure appear "spinless". The band structure has been plotted in the slightly unusual choice of Brillouin zone $[0, 2\pi] \times [0, 2\pi]$, as the Dirac cone here occurs at (π, π). However, the picture is similarly centered around $(0, 0)$ if $\mu = -4$ is chosen instead. Grey plane indicates the Fermi Level.

To get a better idea of the problem, we now plot the band structure for carefully selected parameters. In Fig. 5.5, the four bands are plotted for three different strengths of the Zeeman term. We indeed see that the system exhibits both gapped regimes, as well as a dispersion relation with a Dirac cone, depending on the choice of parameters. Moreover, we see how the top and bottom bands are split off more and more as V_z is increased.

We have now seen both the band structure and the analytical expression for the dispersion relation. However, from inspection of these it is hard to say which parameters lead to a gapped spectrum continuously connected to the vacuum, and which lead to one that needs to go through a gap closing. This is where the Chern number becomes important, because it allows the topological properties to be extracted also in this less obvious case. If the Chern number is non-trivial, we know from Chapter 3 that the gap needs to close at the edge where the bulk meets the vacuum. We will not go into details here, but instead just mention that the Chern number has been calculated for all possible parameters [30]. If we restrict ourselves to the case with $t = 1, \mu = \pm 4$, which is the regime in which we model our systems,[5] then the condition for being in the topologically non-trivial phase is

$$|\Delta| < |V_z| < \sqrt{|\mu|^2 + |\Delta|^2}. \tag{5.24}$$

[5] In a semiconductor the Fermi level is near the band edge. $\mu = \pm 4t$ puts the Fermi level close to the band edge, as can be seen from Eq. (5.12). It is, however, not exactly at the band edge because other terms, most notably the Zeeman term, further shifts the band structure both up and down depending on spin. In the case of an ordinary bulk superconductor, the chemical potential is on the other hand sitting in the middle of the band. However, also in this case a chemical potential sitting at the band edge can be motivated at the surface, because at the surface the work function [73] pushes the bands away from the Fermi level, thereby creating a layer close to the surface where this is true.

For other values of $|V_z|$, the system is in a phase trivially connected to the vacuum. From this relation, and our knowledge about the Chern number from Section 2.4, it is clear why the band structure in Fig. 5.5 has a Dirac cone at $V_z = \Delta$. It is at this point that the system passes between the two phases, and we now know that it is the rightmost band structure, where $V_z = \frac{3\Delta}{2}$, that is topologically disconnected from the vacuum.

5.4.4 Vortices, Majorana fermions, and degenerate ground states

Having seen that the Hamiltonian has a topologically non-trivial phase, we now turn to the Majorana fermions. In analogy with the p-wave superconductor described in Section 5.3, we expect to find Majorana fermions localized at vortex cores. This has analytically been proven to be the case also in this system [25, 30]. The geometry described in Section 5.3.4 is a disk with a hole in, through which a magnetic flux is threaded. The Majorana fermion can therefore be seen as residing on the edge that surrounds the hole. Making the edge very small, it is clear that quantization effects along the edge will make the spectrum there highly discretized. A Majorana fermion at a vortex core can therefore be expected to correspond to an energy level that has a certain excitation gap to the next higher energy state. This means that the ground state of a system with such a Majorana state can be gapped and degenerate. With gapped we here mean both the spectrum in the bulk, and locally at the Majorana state. This is in contrast to the infinite strip, where the bulk is gapped, but a continuous branch cuts through $E = 0$ at the edge, such that the edge is gapless. To understand why the ground state is degenerate we have to consider not only the vortex core, but the whole system. We note, however, that although the discussion below is carried out for a vortex, it is generally valid for any pair of Majorana fermions.

In this chapter we have mainly been concerned with locally non-degenerate solutions to the Bogoliubov-de Gennes Hamiltonian. However, globally we know that each such quasi-particle has a partner with the same energy. Consider now the two Majorana states γ_o and γ_i, on the outer and inner edge of the system. As Majorana fermions they satisfy $\gamma_o^\dagger = \gamma_o$ and $\gamma_i^\dagger = \gamma_i$. Further, because they both have the same energy $E = 0$, it is equaly valid to form two rotated eigenstates by taking the linear combinations

$$\lambda_1 = \frac{1}{\sqrt{2}} \left(\gamma_o + i\gamma_i \right), \tag{5.25}$$

$$\lambda_2 = \frac{1}{\sqrt{2}} \left(\gamma_o - i\gamma_i \right). \tag{5.26}$$

These are eigenstates partly localized at the inner and outer edges. It is also clear that these operators satisfy

$$\lambda_1^\dagger = \frac{1}{\sqrt{2}}\left(\gamma_o^\dagger - i\gamma_i^\dagger\right) = \frac{1}{\sqrt{2}}\left(\gamma_o - i\gamma_i\right) = \lambda_2, \tag{5.27}$$

$$\lambda_2^\dagger = \frac{1}{\sqrt{2}}\left(\gamma_o^\dagger + i\gamma_i^\dagger\right) = \frac{1}{\sqrt{2}}\left(\gamma_o + i\gamma_i\right) = \lambda_1. \tag{5.28}$$

That is, occupying the state associated with λ_1, goes together with deoccupying the state associated with λ_2, and vice versa. Further, because γ_o and γ_i are Majorana fermions, we know that they can be written as $\gamma_o = (c_o + c_o^\dagger)/\sqrt{2}$ and $\gamma_i = (c_i + c_i^\dagger)/\sqrt{2}$, for some electron operators c_o and c_i localized on the outer and inner edge, respectively.[6] It follows that λ_1 and λ_2 also can be written as

$$\lambda_1 = \frac{1}{\sqrt{2}}\left(\eta_1 + \eta_2^\dagger\right), \tag{5.29}$$

$$\lambda_2 = \frac{1}{\sqrt{2}}\left(\eta_2 + \eta_1^\dagger\right), \tag{5.30}$$

where $\eta_1 = (c_o + ic_i)/\sqrt{2}$ and $\eta_2 = (c_o - ic_i)/\sqrt{2}$. Now, η_1 and η_2 corresponds to two different delocalized electron modes. It is therefore clear that $\lambda_1^\dagger = \lambda_2$ is an operator that creates a particle in state η_1, while simultaneously destroying one in η_2, while $\lambda_2^\dagger = \lambda_1$ has the opposite effect.

In Section 4.3.2 we expressed the view that the Bogoliubov operators has to be understood in connection with the BCS wave function as a technical construct that simplifies the mathematics. Further, the Bogoliubov operators, in spite of appearing as a mixture of electron and hole, actually creates an ordinary unpaired electron as an excitation on top of the superconducting condensate. The physical reality therefore seems to me to be the following. The ground state is indifferent to whether the state η_1 or η_2 is occupied, as long as one and only one of them is so. The λ operators can therefore be seen to reshuffle electrons between the two states. The true meaning of a pair of Majorana fermions can therefore be traced back to two ordinary one electron states, which both are partly localized at two different places in space. These two states are fractionally occupied in the sense that only one of the states is occupied at a time. The Majorana fermion picture therefore arise when we choose to look at the problem in a local basis.[7] Moreover, because the ground state is indifferent to which of the two states that is occupied, the ground state is degenerate. Although this means that the Majorana fermion is somewhat of an artefact of

[6] This follows from the fact that we know that γ_o and γ_i are localized on the edges, and that they satisfies the Majorana condition $\gamma^\dagger = \gamma$.

[7] Strictly speaking it is not related to looking at it in a local basis, but rather to looking at it in the basis where $\gamma_o = \gamma_o^\dagger$ and $\gamma_i = \gamma_i^\dagger$, rather than the basis in which $\lambda_1 = \lambda_2^\dagger$. However, this is equivalent to the local basis in all situation of interest to us.

looking at the problem in a local basis, it is not irrelevant. Kitaev's argument is that the non-local nature of the Majorana fermions make them insensitive to local perturbations. The state can only be perturbed through non-local perturbations that act on two Majorana fermions as an ordinary electron, or equivalently, which make the two Majorana fermions interact [34]. According to the view expressed here, such an interaction would mean that the electron is reshuffled among the two available states.

5.5 Non-Abelian statistics and quantum computation

Having understood that Majorana fermions indeed can occur in condensed matter systems, we end this chapter with a few words about why Majorana fermions are interesting. The interest can mainly be divided into two categories. First, Majorana fermions are interesting from a fundamental point of view. As mentioned at the beginning of this chapter, Majorana proposed his equation for fundamental particles. However, no fundamental particles have so far been confirmed to actually be Majorana fermions. Although the Majorana fermions we study here are not fundamental particles, it is still pushing the boundaries of our basic understanding of quantum mechanics. To understand the second reason for why Majorana fermions are of interest, we first need a brief introduction to the field of quantum computation. The presentation here will only cover the minimal amount necessary for understanding the basic idea. For a more thorough introduction to the topic, the reader is referred to [74–76].

5.5.1 Quantum computation

A quantum computer is a proposed computer architecture that utilizes the exponentially large Hilbert space that is required to describe quantum mechanical systems, to effectively perform computational tasks not feasible on ordinary computers. In an ordinary computer, information is stored in terms of ones and zeros called bits, and computations are performed through logic operations on these bits. Let for this reason $[00100110]$ be a state of a binary 8-bit register. An ordinary digital computer carries out computations on this register through successive applications of logic gates. Therefore, let us denote the ordinary NOT gate by N, and assume for simplicity an architecture that can perform one instruction per clock cycle. We can then as an example write down the process of double negation NN of the register as

$$NN[00100110] \rightarrow N[11011001] \rightarrow [00100110], \tag{5.31}$$

where each arrow indicates the passing of one clock-cycle. To perform calculations on multiple data sets, it is further required that data is sequentially loaded into the register and then executed upon. Say for example that we want

to negate [00000000], [00000001] and [00000010]. Letting an arrow followed by a new line indicate loading a new number into the register, this calculation can in line with the customary fetch-execute cycle be written as

$$\begin{aligned} &N[00000000] \rightarrow [11111111] \rightarrow \\ &N[00000001] \rightarrow [11111110] \rightarrow \\ &N[00000010] \rightarrow [11111101]. \end{aligned} \tag{5.32}$$

It is clear that the number of operations, or number of arrows, grows linearly with the amount of data. It is however possible to speed up the calculation by increasing the register size. By doubling the number of bits in the register, two calculations can be performed in parallel, and twice the amount of computations can be performed in the same time. We say that the computational power grows linearly with the size of the system.

Consider now any quantum system that has two possible states, and label these states 0 and 1, respectively, and call this system a qubit. A typical state for a collection of eight qubits can then be written as $|00100110\rangle$, which resembles the ordinary 8-bit digital register. There is one very important difference though. A quantum state for a collection of quantum mechanical qubits can be in a superposition such as

$$|\Psi\rangle = a|00000000\rangle + b|00000001\rangle + c|00000010\rangle. \tag{5.33}$$

A quantum mechanical NOT gate N can then be applied to this state in a single cycle, making it possible to perform the triple negation above in a single instruction

$$N|\Psi\rangle \rightarrow a|11111111\rangle + b|11111110\rangle + c|11111101\rangle. \tag{5.34}$$

Further, the total number of states $|\Psi\rangle$ can be in a superposition of is 2^8. In principle it is therefore possible to perform 2^8 instructions in a single cycle on a quantum computer that implements such a scheme. In general a system consisting of N qubits can be in a superposition of 2^N states, and can therefore be used to carry out up to 2^N instructions per cycle. The number of possible calculations grows exponentially with the system size. Flipping ones and zeros is, however, not the only computational task that can be performed on quantum states. The coefficients a, b and c in the example above are complex numbers, and it turns out that it is also important in quantum computation to have detailed control of these [75].

Although the above arguments makes quantum computing look tremendously attractive, in practice there are many obstacles to a working quantum computer. First of all quantum mechanics implies that if we try to read out any single one of the results of $N|\Psi\rangle$, the wave function will collapse onto a single state. This means that although an exponential number of calculations can be performed at once, if a naive readout process is chosen, then only a single

of those results will be available to us. In spite of this, algorithms exists that are able to utilize the exponential power of quantum computers, should they become available [76]. A problem more relevant to us is, however, the problem of decoherence [74, 76]. Decoherence is the process whereby a quantum mechanical system loses phase coherence due to interaction with the environment. This is a serious problem for the implementation of quantum computers, and the main obstacle for scaling up quantum computers from any more than a few qubits. The main motivation behind the idea of topological quantum computing is exactly to get around the problem of decoherence.

5.5.2 Topological quantum computation

We are now ready to understand the second and more practical reason for why Majorana fermions are of interest. Kitaev in 2001 proposed the use of Majorana modes at the end points of superconducting wires, to robustly encode quantum information in a way that protects it from decoherence [34]. However, that the Majorana fermions are located at the endpoints of wires is irrelevant, and for example Majorana fermions located at vortex cores can be used for the same purpose. In particular, the delocalized nature of the underlying electron, as well as the observation in Section 5.4.4 that Majorana fermions in vortex cores have an excitation gap is important. These properties are true also for Majorana fermions at wire end points. According to Kitaev's argument, the delocalized nature of the underlying electron prevents it from being perturbed by local interactions. At the same time the energy gap is important because it also prevents the individual Majorana fermions from being locally excited, as well as other low lying excitations from interfering with the Majorana fermions during meassurements [77].

It has also been shown that under the interchange of two vortices, say clockwise, the Majorana fermions at vortex i and j transform according to[8] [78]

$$\gamma_i \to \gamma_j, \tag{5.35}$$

$$\gamma_j \to -\gamma_i. \tag{5.36}$$

Interchanging the two vortices once more, we complete a full revolution of the vortices around each other, resulting in both γ's having acquired a minus sign. This is an example of the anyonic statistics that fermions can have under interchange in two dimensions [79]. However, it turns out that the transformation is not only anyonic. Under the interchange of multiple vortices, so called braiding, the signs acquired will depend on the order of interchange. The statistics is therefore non-Abelian [78]. We can trace this behavior back to the Majorana operators. We know from the previous sections that Majorana

[8]Here i and j labels the position of the vortex rather than the vortex itself. This is why the operators change index, as they continuously are brought from point i to j following along with the vortices.

fermions are local parts of a delocalized electron. As vortices are braided, half electrons will move around each other, and the statistics for this turns out to be non-Abelian. The non-Abelian statistics can further be utilized for topological quantum computation, as it provides a way through which the signs in front of the different basis states can be manipulated. Majorana fermions do therefore not only present a way for decoherence free storage of quantum information, but also a way to perform computational operations on that information [78]. Finally, we mention that a scheme for braiding Majorana fermions situated at the endpoints of one-dimensional wires also has been proposed [80].

Part III:
Method development

6. Tight-Binding ToolKit (TBTK)

While the theoretical background in the previous part is described in analytical terms, most of the results in this thesis rely heavily on numerical calculations. As a result, a numerical library for solving general bilinear Hamiltonians has evolved in parallel with the results detailed in the Results Part and Papers. At the time of writing, the library consists of more than 15,000 lines of code,[1] is available for download online,[2] and can solve general bilinear models using either wave function or Green's function methods. We here give a basic introduction to the library, with the purpose of conveying the core design philosophy. This is done through a discussion of the main features, together with code snippets that demonstrate the workflow. For clarity we point out that these code snippets are not standalone programs ready to be compiled. For such examples we instead refer to the templates that come packed together with the library. Another useful resource for a deeper understanding of the library is the application programming interface (API).[3]

6.1 Modelling

6.1.1 Bilinear Hamiltonian and physical indices

TBTK is designed with the purpose of modelling and solving arbitrary bilinear Hamiltonians, which can be written as

$$H = \sum_{\mathbf{ij}} a_{\mathbf{ij}} c^\dagger_{\mathbf{i}} c_{\mathbf{j}}. \tag{6.1}$$

[1] The main part of the library, which is used for calculations, is written in c++11. Part of the library is also parallelized using CUDA (`http://www.nvidia.com/object/cuda_home_new.html`) to run on one or multiple GPUs, but can be installed to run also on machines without a GPU. Further, Lapack (`http://www.netlib.org/lapack/`) and Blas (`http://www.netlib.org/blas/`) need to be installed on the computer, while a local copy of HDF5 (`https://www.hdfgroup.org/HDF5/`) is downloaded and compiled as part of the installation procedure. Basic plotting scripts for common physical quantities are also provided and are implemented using python. While the code has been tested on PC with Linux (Ubuntu) and Mac, and with a few different compiler combinations, most extensive testing have been done using Scientific Linux release 6.7, gcc 4.9, and NVCC 7.0. Most importantly, a gcc compiler supporting c++11 has to be used.

[2] `http://dafer45.github.io/TBTK/`

[3] `http://dafer45.github.io/TBTK/doc/html/index.html`

Here **i** and **j** are arbitrary indices, and an important feature of TBTK is the way in which it allows the user to model complex systems using physical indices.

To understand how TBTK handles indices, we make the following distinction between physical and Hilbert space indices. A physical index is an index such as (x, y, s), where x and y are spatial coordinates, and s is a spin index. For simplicity, assume that x and y are indices on a square lattice with size SIZE_X$\times$SIZE_Y. In order to write the Hamiltonian on a matrix form, the physical indices have to be converted into one-dimensional Hilbert space indices, which e.g. can take the form $h = 2 \times \text{SIZE\_Y} \times x + 2 \times y + s$. However, the exact form of the mapping is dependent on the system of interest. Explicitly handling the Hilbert space indices is a nuisance when modeling a system, and obscures the physics in the code. To solve this TBTK allows the user to always work with physical indices, converting these on the fly to Hilbert space indices in the underlying algorithms that are more naturally expressed in Hilbert space indices.[4]

In addition to taking the burden off the user from having to handle Hilbert space indices, TBTK simultaneously allows for multiple physical indexing schemes to be employed in a single model. Let us for example consider a system consisting of a three-dimensional cubic lattice, connected to a two-dimensional hexagonal lattice through a one-dimensional molecule. Further, assume that that the three-dimensional bulk requires an additional orbital index, while the hexagonal sheet requires a sublattice index. The three subsystems are then naturally described using the three indexing schemes (x, y, z, o, s), (x, y, l, s), and (x, s), respectively, where x, y, and z are spatial coordinates, while o, l, and s are orbit, sublattice, and spin indices, respectively. We note that although for example x is used in all three indexing schemes, there is no relation between the x's in the different schemes, and they can all have different ranges. TBTK allows for such different indexing schemes to be employed by simply prepending the indices by yet another subsystem index.[5] In fact, TBTK neither requires that a particular subindex take on every value in a range. It is therefore perfectly valid to exclude a site in the middle of a lattice to model a vacancy. The Hilbert space that is constructed by TBTK will only map to indices actually included in the model, and will therefore ensure that a minimal Hamiltonian always is constructed.

[4] Technically one of the corner stones of TBTK is a custom Tree class called TreeNode that acts both as storage for the hopping amplitudes $a_{\mathbf{ij}}$ and a dictionary between physical and Hilbert space indices. Due to its tree structure, it allows for both hopping amplitudes to be retrieved, and indices to be converted, with minimal performance penalties independently of the size of the Hilbert space. The tree structure also has the additional benefit of naturally enabling multiple indexing schemes to be used in a single model.

[5] In general, if two indices differ in value for a particular subindex, then the structure of the subindices to the right of that subindex can differ in structure. Indices that separates structurally different components of a system should therefore be put as far to the left as possible.

6.1.2 Modeling superconductivity

The Hamiltonian in Eq. (6.1) is a completely general bilinear Hamiltonian, and as such can describe any single particle (non-interacting) problem with discrete indices. As TBTK is constructed to model and solve any such problem,[6] this means that any non-interacting system with discrete indices can be modelled and solved. Moreover, in Section 4.2.2 we noted that, in spite of being interacting, the Bogoliubov-de Gennes mean field treatment reduced also the superconducting problem to a bilinear form. However, in contrast to Eq. (6.1), the superconducting problem also has annihilation operators to the left, and creation operators to the right. This is easily solved by letting the spin index run over four values instead of two. That is, by interpreting the Hamiltonian in terms of a Nambu-basis, where the two first components correspond to spin up and spin down electrons, while the last two correspond to their hole counter parts.

6.1.3 Example

For a basic demonstration of how to use the TBTK library to setup a model, we consider the Hamiltonian

$$H = -t \sum_{\langle \mathbf{ij} \rangle \sigma} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{j}\sigma} - \mu \sum_{\mathbf{i}\sigma} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{i}\sigma}, \tag{6.2}$$

where **i** and **j** are indices on a two-dimensional lattice of size 20×20, σ is a spin-index, $t = 1$, and $\mu = -2$. The model is created using

```
//Lattice size
const int SIZE_X = 20;
const int SIZE_Y = 20;

//Parameters
complex<double> t = 1.0;
complex<double> mu = -2.0;

//Create model and add hopping amplitudes
Model model;
for(int x = 0; x < SIZE_X; x++){
  for(int y = 0; y < SIZE_Y; y++){
    for(int s = 0; s < 2; s++){
      //Add hopping amplitudes corresponding to chemical
      //potential
      model.addHA(
        HoppingAmplitude(
          -mu,
```

[6] Within the limitations of available computational resources, currently tractable problems are those that have a Hilbert space size of up to around 10^4 for the wave function method (diagonalization), and 10^6 for the Green's function method.

```
            {x, y, s},
            {x, y, s}
          )
        );

        //Add hopping amplitudes corresponding to t
        if(x+1 < SIZE_X)
          model.addHAAndHC(
            HoppingAmplitude(
              -t,
              {(x+1)%SIZE_X, y, s},
              {x, y, s}
            )
          );
        if(y+1 < SIZE_Y)
          model.addHAAndHC(
            HoppingAmplitude(
              -t,
              {x, (y+1)%SIZE_Y, s},
              {x, y, s}
            )
          );
      }
    }
  }

  //Construct Hilbert space
  model.construct();
```

The code starts by listing the relevant parameters. Then a Model[7] is created and the indices of the model are looped over. Inside the loop body, the first line adds a HoppingAmplitude to the Model, which should be understood to correspond to the diagonal matrix elements $a_{\mathbf{ii}} = -\mu$. Next, the HoppingAmplitudes corresponding to $a_{\mathbf{ij}} = -t$ are added, and in particular the function 'addHAAndHC' is used to simultaneously add the HoppingAmplitude and its Hermitian conjugate. The if-statements guard against the addition of a HoppingAmplitude at the very end of the lattice, and thereby implements open boundary conditions. However, if the if-statements are commented out, the code is prepared to result in periodic boundary conditions, as the modulo sign (%) clamps the indices to the range $[0, \text{SIZE\_X} - 1]$ and $[0, \text{SIZE\_Y} - 1]$, respectively. Finally, in the last line a Hilbert space basis is constructed by creating a one-to-one mapping between the physical indices fed into the Model, and a monotonously increasing natural number. The Model is now ready to be solved.

[7] We use upper case letters to refer to the classes and objects in the code, to distinguish them from the same words without such a specific technical meaning. In particular, in this sentence 'Model' refers to the object model in the code, while the subsequent use of the word 'model' refers to the mathematical model in a broader sense.

6.1.4 Additional information

The Model is not only a container of the Hamiltonian, but all model related information. As such it is also possible to set the temperature, chemical potential, and statistics of the model using

```
//Set temperature
model.setTemperature(293);

//Set chemical potential
model.setChemicalPotential(-1);

//Set statistics
model.setStatistics(Model::Statistics::BoseEinstein);
```

These parameters are used in the property extractors described in Section 6.2.2. By default they are set to 0, 0, and 'Model::Statistics::FermiDirac', respectively. We note that if the chemical potential is set in this way, it should not be included in the Hamiltonian.

6.2 Solving

6.2.1 Solvers

As mentioned in the introduction to this Chapter, TBTK can be used to solve Models using either wave function methods, or Green's function methods. This is accomplished through the use of different solvers, and currently TBTK has two types of solvers: a DiagonalizationSolver, and a ChebyshevSolver. The former diagonalizes the Hamiltonian and thereby results in a solution to the problem in terms of eigenvectors and eigenvalues. That is, it calculates the wave functions $\Psi_n(\mathbf{i})$ and energies E_n, respectively. The second method calculates the Green's function $G_{\mathbf{ij}}(E)$ using a Chebyshev expansion, where the Chebyshev expansion itself is described in more detail in Chapter 7.

As a complete knowledge of either the wave function and energies, or the Green's function, is enough to determine all properties of a model, only one method is needed in principle. However, computationally as well as conceptually the two methods have different strengths and therefore complement each other. One of the main advantages of the ChebyshevSolver is that it can handle much larger problems, because for many quantities the computation time scales as $O(n)$ with the size n of the Hilbert space. In contrast, the computation time for the DiagonalizationSolver always scales as $O(n^3)$. In return, the DiagonalizationSolver for example gives direct access to the wavefunction, which can provide valuable physical insight into the problem.

To set up and run a DiagonalizationSolver, the following code is used

```
DiagonalizationSolver dSolver;
dSolver.setModel(&model);
dSolver.run();
```

Similarly, a ChebyshevSolver is setup using

```
ChebyshevSolver cSolver;
cSolver.setModel(&model);
cSolver.setScaleFactor(SCALE_FACTOR);
```

The first two lines which setup the solvers are conceptually identical. This is intentional, with the purpose of making it as simple as possible to switch between the use of different solvers. In spite of this, the DiagonalizationSolver and ChebyshevSolver are different and require somewhat different workflows when handled directly. Most notably, due to the inner workings of the algorithms, the DiagonalizationSolver needs to diagonalize the whole problem before properties can be calculated, while the ChebyshevSolver calculates properties on the fly. This is reflected in the call to 'dSolver.run()' above, which diagonalizes the problem and puts it into a state where it is ready to be used to extract properties.[8] In contrast, the ChebyshevSolver is in this state immediately after the Model has been set and only a scale factor needs to be set. The reason why a scale factor is needed is explained in Chapter 7.

6.2.2 Property extractors

In the previous section we saw that solvers are set up in a very similar way, but also learned that they require different workflows. To further alleviate this problem, and to provide methods for extracting standard properties from the model, TBTK provides property extractors. These are the preferred objects to use for extracting any properties, and direct use of the solvers is therefore discouraged. A property extractor for a DiagonalizationSolver is created using

```
DPropertyExtractor pe(&dSolver);
```

Similarly, a property extractor for a ChebyshevSolver is created using

```
CPropertyExtractor pe(
  &cSolver,
  NUM_COEFFICIENTS,
  ENERGY_RESOLUTION,
  false,
  false,
  true,
  LOWER_LIMIT,
  UPPER_LIMIT
);
```

We note that although both property extractors wrap a solver, they require different sets of parameters. This is reflective of the fundamental differences that exist between the solvers. A basic understanding of the methods the solvers

[8] The reason the method is called 'run()' rather than 'diagonalize()' is that it actually will run a full self-consistent loop if the DiagonalizationSolver and model is appropriately configured before the call. For further details on this we refer to the templates in the TBTK package.

are based on is therefore needed to successfully set up the property extractors. However, once the property extractors are set up, the rest works the same. Diagonalization should be a well-known method, and we therefore give no further explanation here. The two parameters that follow the cSolver will become clear in Chapter 7, while the two last parameters are the energies between which the relevant Green's functions are calculated. The three Boolean options have to do with specifying whether or not specific parts of the calculation should be done on a GPU or not, and whether to use a lookup table to speed up calculations. For more information about these we refer to the API.

Having wrapped the solvers in property extractors, we can now extract a number of different properties using

```
//Calculate density
Property::Density *density = pe.calculateDensity(
  {IDX_X, IDX_Y, IDX_SUM_ALL},
  {SIZE_X, SIZE_Y, 2}
);

//Calculate magnetization
Property::Magnetization *mag = pe.calculateMagnetization(
  {IDX_X, IDX_Y, IDX_SPIN},
  {SIZE_X, SIZE_Y, 2}
);

//Calculate spin-polarized local density of states (LDOS)
Property::SpinPolarizedLDOS *spLdos =
  pe.calculateSpinPolarizedLDOS(
    {IDX_X, SIZE_Y/2, IDX_SPIN},
    {SIZE_X, 1, 2},
    LOWER_LIMIT,
    UPPER_LIMIT,
    RESOLUTION
  );
```

We do not go into details about the arguments or format of the return values here, but rather list the commands with the purpose of demonstrating the workflow. The interested reader is referred to the API and the templates for more details.

In addition to the common properties, there are some properties that for practical reasons only can be extracted from one or the other of the solvers. The DPropertyExtractor can extract the following additional properties

```
//Calculate density of states (DOS)
Property::DOS *dos = pe.calculateDOS(
  LOWER_LIMIT,
  UPPER_LIMIT,
  RESOLUTION
);

//Get eigenvalue for state n
```

```
double E = pe.getEigenValue(n);

//Get all eigenvalues
double *allE = pe.getEigenValues();

//Get amplitude for index {x, y, s} and eigenstate n
complex<double> amplitude = pe.getAmplitude(
  n,
  {x, y, s}
);
```

In contrast, the CPropertyExtractor can extract arbitrary two operator expectation values and the Green's function

```
//Calculate <c_{i}^{\dagger}c_{j}>, where i = (x, y, s)
//and j = (x+1, y, s)
complex<double> expectationValue =
  pe.calculateExpectationValue(
    {x, y, s},
    {x+1, y, s}
  );

//Calculate G_{ij}(E), where i = (x,y,s) and j = (x+1,y,s)
complex<double> *g = pe.calculateGreensFunction(
  {x, y, s},
  {x+1, y, s},
  TYPE
);
```

where

$$\texttt{TYPE} \in \left\{ \begin{array}{c} \texttt{ChebyshevSolver::GreensFunctionType::Retarded} \\ \texttt{ChebyshevSolver::GreensFunctionType::Advanced} \\ \texttt{ChebyshevSolver::GreensFunctionType::Principal} \\ \texttt{ChebyshevSolver::GreensFunctionType::NonPrincipal} \end{array} \right\}.$$

More information about these different types of Green's functions is provided in Chapter 7.

6.3 File writer and plotting scripts

While we have given no detailed description of the format of the data returned by the property extractors here, we mentioned that it is returned in a format that is standardized across TBTK. What this means is that not only do the property extractors output the information in a common format, but also other parts of the code understand it. Namely, the FileWriter and plotting scripts. Data extracted by a property extractor can therefore be written immediately to file, and then be plotted with the predefined plotting scripts. Here we give an

example of how this is done, and refer the reader to the API and templates for further information and examples.

Before writing values to file, it is good practice to set the filename[9] and clear any possible copy of the file from the working directory. This is preferably done once at the start of the program using

```
FileWriter::setFileName("MyResults.h5");
FileWriter::clear();
```

We note the use of the file extension .h5, which is a standardized scientific file format used by TBTK.[10] As such, the data can be imported into most numerical programming environments for post-processing and plotting, e.g. MATLAB, Mathematica, Python, or Julia. To write the DOS calculated above, we can now pass it to the FileWriter using

```
FileWriter::writeDOS(dos);
```

Having written the DOS to file, it is now possible to plot the data from the terminal using

```
python ${TBTK_dir}/TBTK/view/python/plotDOS.py MyResults.h5 0.1
```

In this case the last parameter is a smoothing parameter giving the delta peaks in the DOS a Gaussian broadening with $\sigma = 0.1$.

6.4 Unit handler

The discussion has so far been carried out in arbitrary units. However, TBTK provides a UnitHandler to simplify the connection to experimentally relevant units. The UnitHandler is based on a principle similar to the SI unit system,[11] where a distinction is made between base units and derived units. However, while the choice of base units have considerable overlap, they are not identical. Rather, the choice of base units has been made in order to be more relevant to quantum mechanical calculations. Moreover, unlike the SI system, the base unit for e.g. length is not required to be m, but rather a number of different units such as m, mm, nm, and $\AA$ can serve as base unit. For a complete list of available base units we refer to the API, and here only mention that the quantities for which base units are defined are temperature, time, length, energy, charge, and count.[12] The default base units are Kelvin (K), seconds (s), meter (m), electron volt (eV), Coulomb (C), and pieces (pcs). The quantities for which there exist special support in terms of derived units are currently mass and magnetism, e.g. kilogram (kg) and Tesla (T), respectively. However, we will not go into details about the use of derived units here.

[9] The default filename is TBTKResults.h5

[10] https://www.hdfgroup.org/HDF5/

[11] http://physics.nist.gov/Pubs/SP330/sp330.pdf

[12] Count refers to numerical quantities measured in pcs or mol.

In addition to base units and derived units, the UnitHandler also supports the use of natural units, which are in one-to-one correspondence with the base units. The purpose of these is to allow for simple numbers such as $t = 1$ to be used in calculations, even if the hopping parameter actually is say $274.19meV$. All function calls in the library should be understood in terms of natural units, except for the calls that set the natural units themselves. Further, TBTK will on the fly convert these values to base units whenever they need to be multiplied by physical constants, such as for example when evaluating the Fermi-Dirac distribution $1/(e^{(E-\mu)/kT} + 1)$.

The UnitHandler is preferably initialized at the very start of the program, for example by specifying that the energy base unit is meV through the call

```
UnitHandler::setEnergyUnit(UnitHandler::EnergyUnit::meV);
```

The natural energy scale can then be set to $274.19meV$ through the subsequent call

```
UnitHandler::setEnergyScale(274.19);
```

Similar pairs of calls can be made for the five remaining quantities. Although it is good practice to specify the base and natural units explicitly like this, we note that it is often not needed because most of them are irrelevant in basic calculations. The units that most commonly need to be specified are temperature and energy, because they enter into the Fermi-Dirac distribution and become important at non-zero temperatures.

7. Chebyshev expansion

In the previous chapter we mentioned that both wave function and Green's function methods are implemented in TBTK through the DiagonalizationSolver and ChebyshevSolver. While diagonalization can be considered a standard procedure and requires little explanation, the Chebyshev expansion of the Green's function cannot. In this chapter we therefore give an introduction to the Chebyshev expansion, deriving it from the Fourier expansion, and applying it to the Green's function. The result is an expression that can be used to recursively calculate the expansion coefficients of the Chebyshev expansion in a way that requires only sparse matrix-vector multiplications. Further, GPUs are specifically designed with the purpose of handling matrix operations, and this is exploited in TBTK. This is one of the reasons why the ChebyshevSolver is able to handle models with comparatively large Hilbert spaces. For more background on the Chebyshev expansion and its applications to the Green's function, we refer to the references [81–83].

We note that because this chapter is mainly concerned with the derivation of the Chebyshev expansion, which is a purely mathematical concept, it is notably more technical than the other chapters. For the reader that simply wants a rough understanding of what we mean by a Chebyshev expansion of the Green's function in the previous chapters, we therefore mention that it is a function expansion of the form

$$G_{\mathbf{ij}}(E) = \frac{2i}{\sqrt{s^2 - E^2}} \sum_{m=0}^{\infty} \frac{b_{\mathbf{ij}}^{(m)}}{1 + \delta_{0m}} \cos(m \operatorname{acos}(E/s)). \tag{7.1}$$

Here $\cos(m \operatorname{acos}(E/s))$ in fact also can be $i \sin(m \operatorname{acos}(E/s))$ or $e^{\pm im \operatorname{acos}(E/s)}$, depending on which particular Green's function that is expanded. Further, the ChebyshevSolver is executed in two steps, where the first calculates the expansion coefficients $b_{\mathbf{ij}}^{(m)}$, and the second evaluates the expression for the Green's function above for a given number of energy values.

7.1 Expanding a function

7.1.1 Fourier expansion in disguise

We begin by deriving the Chebyshev transform and Chebyshev expansion from the corresponding Fourier cosine transform and Fourier cosine expansion.

We will see that given a function $f(x)$ defined on the interval $x \in [-1, 1]$, the Chebyshev transform (expansion) in x can be thought of as the Fourier cosine transform (expansion) of $\tilde{f}(\theta) = f(\text{acos}(\theta))$, in the variable θ. This leads to a set of Chebyshev basis functions $T_m(x) = \cos(m \,\text{acos}(x))$, in correspondence with the Fourier basis functions $\cos(m\theta)$. Further, we show that $T_m(x)$ are polynomials of degree m, and that they satisfy a recursion relation.

In more detail $T_m(x)$ are called Chebyshev polynomials of the first kind, and we refer to the corresponding Chebyshev transform (expansion) as the Chebyshev transform (expansion) of the first kind. It is also shown that a closely related set of basis functions $U_m(x)$ known as Chebyshev polynomials of the second kind can be obtained starting from the Fourier sine transform (expansion) of $\tilde{g}(\theta) = \sin(\theta)\tilde{f}(\theta)$. The corresponding Chebyshev transform (expansion) is here referred to as the Chebyshev transform (expansion) of the second kind.

Finally, it is shown that a third hybrid version of the Chebyshev transform (expansion) of the first and second kind leads to what we here refer to as a Chebyshev transform (expansion) of the alternative kind. In particular, it is this transform (expansion) that is applied to the Green's function in Section 7.2.

7.1.2 First kind

Chebyshev transform of the first kind

Consider the Fourier cosine transform of the symmetric function $\tilde{f}(\theta)$

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} \tilde{f}(\theta)\cos(m\theta)d\theta = \frac{2}{\pi}\int_{-\pi}^{0} \tilde{f}(\theta)\cos(m\theta)d\theta. \tag{7.2}$$

Next perform the change of variables $\theta = \text{acos}(x)$, such that

$$a_m = \frac{2}{\pi}\int_{-1}^{1} \tilde{f}(\text{acos}(x))\cos(m\,\text{acos}(x))\frac{dx}{-\sin(\text{acos}(x))}. \tag{7.3}$$

Now noting that $\sin(\text{acos}(x))$ is negative on the interval $\theta \in [-\pi, 0]$, we can write $-\sin(\text{acos}(x)) = \sqrt{1-\cos^2(\text{acos}(x))} = \sqrt{1-x^2}$. Further, defining $f(x) = \tilde{f}(\text{acos}(x))$, we arrive at

$$a_m = \frac{2}{\pi}\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}\cos(m\,\text{acos}(x))dx. \tag{7.4}$$

This is known as a Chebyshev transform of $f(x)$. Because $\tilde{f}(\theta)$ can be an arbitrary function on the interval $\theta \in [-\pi, 0]$, also $f(x)$ can be an arbitrary function on the interval $x \in [-1, 1]$.[1] The functions

$$T_m(x) = \cos(m\,\text{acos}(x)) \tag{7.5}$$

are known as Chebyshev polynomials of the first kind.

[1] L^2-integrability assumed.

Chebyshev polynomials of the first kind
To see that $T_m(x)$ really are polynomials, we note that for $m = 0, 1$ we have

$$\begin{aligned} T_0(x) =& 1, \\ T_1(x) =& x. \end{aligned} \tag{7.6}$$

Further, it is shown in Eq. (12.1) in the Appendix that for $m > 1$

$$T_m(x) = 2xT_{m-1}(x) - T_{m-2}(x). \tag{7.7}$$

Together with the initial expressions in Eq. (7.6), this shows that every $T_m(x)$ indeed is a polynomial of degree m.

Chebyshev expansion of the first kind
Knowing that a_m are the Fourier coefficients of $\tilde{f}(\theta)$, we can expand it as a Fourier cosine series

$$\tilde{f}(\theta) = \sum_{m=0}^{\infty} \frac{a_m}{1 + \delta_{0m}} \cos(m\theta). \tag{7.8}$$

From this it follows by variable substitution that

$$f(x) = \sum_{m=0}^{\infty} \frac{a_m}{1 + \delta_{0m}} T_m(x), \tag{7.9}$$

which is the corresponding Chebyshev expansion of $f(x)$.

7.1.3 Second kind

Chebyshev transform of the second kind
We next consider the function

$$\tilde{g}(\theta) = \sin(\theta)\tilde{f}(\theta). \tag{7.10}$$

Knowing that a_m are the Fourier coefficients of $\tilde{f}(\theta)$, we can write

$$\tilde{g}(\theta) = \sum_{m=0}^{\infty} a_m \cos(m\theta) \sin(\theta). \tag{7.11}$$

Next, the coefficients of the Fourier sine transform of $\tilde{g}(\theta)$ can be written as

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} \sum_{m=0}^{\infty} a_m \cos(m\theta) \sin(\theta) \sin(n\theta) d\theta, \tag{7.12}$$

where b_0 is trivially zero, because $\tilde{g}(\theta)$ is anti-symmetric, and we therefore restrict our attention to $n > 0$ below. Now using that

$$\sin((m+1)\theta) = \cos(m\theta)\sin(\theta) + \sin(m\theta)\cos(\theta), \tag{7.13}$$

we have

$$\begin{aligned} b_n =& \frac{1}{\pi}\int_{-\pi}^{\pi} \sum_{m=0}^{\infty} a_m \left(\sin((m+1)\theta) - \sin(m\theta)\cos(\theta)\right)\sin(n\theta)d\theta \\ =& a_{n-1} - \frac{1}{\pi}\sum_{m=0}^{\infty} a_m \int_{-\pi}^{\pi} \sin(m\theta)\sin(n\theta)\cos(\theta)d\theta. \end{aligned} \tag{7.14}$$

We now want to show that the integrals in the second term are identically zero. For this reason we split the integral in two, one part on the interval $\theta \in [-\pi, 0]$, and one part on the interval $\theta \in [0, \pi]$. Next we perform a change of variable $\theta \to \theta - \pi$ on the second interval, resulting in the expression

$$\begin{aligned} &\int_{-\pi}^{\pi} \sin(m\theta)\sin(n\theta)\cos(\theta)d\theta \\ &= \int_{-\pi}^{0} \left(1 - (-1)^{m+n}\right)\sin(m\theta)\sin(n\theta)\cos(\theta)d\theta \end{aligned} \tag{7.15}$$

We now split the integral again onto the intervals $\theta \in [-\pi, -\pi/2]$ and $\theta \in [-\pi/2, 0]$, and perform the change of variable $\theta \to -\theta - \pi$ on the second interval. This results in

$$\begin{aligned} &\int_{-\pi}^{0} \left(1 - (-1)^{m+n}\right)\sin(m\theta)\sin(n\theta)\cos(\theta)d\theta \\ &= \left(1 - (-1)^{m+n}\right)\left(1 + (-1)^{m+n}\right)\int_{-\pi}^{-\pi/2} \sin(m\theta)\sin(n\theta)\cos(\theta)d\theta. \end{aligned} \tag{7.16}$$

Because m and n are integers, either the first or second parenthesis is zero. It follows from Eq. (7.14) that the relation

$$b_n = a_{n-1} \tag{7.17}$$

holds between the Fourier cosine transform of $\tilde{f}(\theta)$, and the Fourier sine transform of $\tilde{g}(\theta)$.

Having derived a relation between a_{m-1} and b_m, we now look at Eq. (7.12), which also can be written as

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} \tilde{g}(\theta)\sin((m+1)\theta)d\theta = \frac{2}{\pi}\int_{-\pi}^{0} \tilde{g}(\theta)\sin((m+1)\theta)d\theta, \tag{7.18}$$

where in the last step the antisymmetry of $\tilde{g}(\theta)$ and $\sin((m+1)\theta)$ have been used. Changing to the coordinates $x = \cos(\theta)$ and denoting $g(x) = \tilde{g}(\text{acos}(x)) = \sin(\text{acos}(x))f(x)$, we have

$$\begin{aligned} a_m =& -\frac{2}{\pi}\int_{-1}^{1} \sin(\text{acos}(x))f(x)\frac{\sin((m+1)\text{acos}(x))}{\sin(\text{acos}(x))}dx \\ =& \frac{2}{\pi}\int_{-1}^{1} \sqrt{1-x^2}f(x)\frac{\sin((m+1)\text{acos}(x))}{\sin(\text{acos}(x))}dx. \end{aligned} \tag{7.19}$$

This expression is known as a Chebyshev transform of the second kind, where the functions

$$U_m(x) = \frac{\sin((m+1)\text{acos}(x))}{\sin(\text{acos}(x))} \tag{7.20}$$

are known as Chebyshev polynomials of the second kind. Notably, it is possible to calculate the Chebyshev (and Fourier) coefficients with either of the two methods

$$a_m = \frac{2}{\pi}\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}T_m(x)dx = \frac{2}{\pi}\int_{-1}^{1} \sqrt{1-x^2}f(x)U_m(x)dx. \tag{7.21}$$

Chebyshev polynomials of the second kind

Showing that the Chebyshev polynomials of the second kind really are polynomials proceeds similarly as for the Chebyshev polynomial of the first kind by noting that for $m = 0$ and $m = 1$ we have

$$\begin{aligned} U_0(x) =& 1, \\ U_1(x) =& 2x, \end{aligned} \tag{7.22}$$

where the second equation follows from $\sin(2\theta) = 2\cos(\theta)\sin(\theta)$. By a similar repeated application of the trigonometric identities as in (12.1), it is also possible to arrive at the recursive relation

$$U_m(x) = 2xU_{m-1}(x) - U_{m-2}(x), \tag{7.23}$$

for $m > 1$. Therefore, also the mth Chebyshev polynomial of the second kind is a polynomial of mth order.

Chebyshev expansion of the second kind

From the fact that $b_m = a_{m-1}$ are the expansion coefficients of $\tilde{g}(\theta) = \sin(\theta)\tilde{f}(\theta)$, it follows that the corresponding Fourier sine expansion of $\tilde{g}(\theta)$ leads to

$$\tilde{f}(\theta) = \frac{1}{\sin(\theta)}\sum_{m=1}^{\infty} b_m \sin(m\theta) = \sum_{m=0}^{\infty} a_m \frac{\sin((m+1)\theta)}{\sin(\theta)}. \tag{7.24}$$

By a variable substitution it therefore follows that

$$f(x) = \sum_{m=0}^{\infty} a_m U_m(x), \tag{7.25}$$

which is known as a Chebyshev expansion of the second kind of $f(x)$.

7.1.4 Alternative kind

In close connection to the Chebyshev expansion of the first and second kind, we will for the Green's function utilize the closely related hybrid expansion

$$f(x) = \frac{2}{\pi\sqrt{1-x^2}} \sum_{m=0}^{\infty} \frac{\mu_m}{1+\delta_{0m}} T_m(x). \tag{7.26}$$

Consider therefore the expression

$$\int_{-1}^{1} f(x) T_n(x) dx = \frac{2}{\pi} \int_{-1}^{1} \sum_{m=0}^{\infty} \frac{\mu_m}{1+\delta_{0m}} T_m(x) \frac{T_n(x)}{\sqrt{1-x^2}} dx = \mu_n, \tag{7.27}$$

where the last equality follows from the fact that this is a Chebyshev transform of the first kind of the Chebyshev expansion of the first kind. The expression

$$\mu_n = \int_{-1}^{1} f(x) T_n(x) dx, \tag{7.28}$$

therefore is the corresponding alternative Chebyshev transform.

7.2 Expanding the Green's function

7.2.1 Analytic and matrix forms

We now derive the Chebyshev transform (expansion) of the alternative kind for the Green's function, which we from now on simply will refer to as the Chebyshev transform (expansion). This is done by first deriving the retarded and advanced Green's functions $G^R_{\mathbf{ij}}(E)$ and $G^A_{\mathbf{ij}}(E)$ in an analytical form, which in turn are split into a "principal" and a "non-principal" Green's function $G^P_{\mathbf{ij}}(E)$ and $G^{NP}_{\mathbf{ij}}(E)$. Next, a Chebyshev transform is performed on $G^{NP}_{\mathbf{ij}}(E)$, followed by the corresponding Chebyshev expansion. This leads to an expansion of the "non-principal" Green's function in terms of $\cos(m\,\mathrm{acos(x)})$ basis functions. It is then shown that the "principal" Green's function is obtained by simply replacing these basis functions by $i\sin(m\,\mathrm{acos}(\theta))$.

The retarded and advanced Green's functions are in turn obtained as superpositions of the "principal" and "non-principal" Green's functions. We also

explain that while the retarded and advanced Green's functions are standard expressions, it is actually the "non-principal" Green's function that is of foremost interest. In analytical calculations the non-principal part is often indirectly isolated by performing one of several tricks from complex analysis, such as integration from $-\infty$ to ∞, integration along a contour, or in special cases by taking the imaginary part of the Green's function. However, here we work with the non-principal part immediately.

The first derivation of the Chebyshev transform (expansion) relies on the analytical expression for the Green's function being known. However, this is not the case in actual calculations as they are performed in TBTK, where only the Hamiltonian is known. For this reason a method for obtaining the Chebyshev expansion coefficients for the Green's function from the Hamiltonian is derived in the final part. In particular, this leads to an algorithm for finding the expansion coefficients that utilizes the recursive nature of the Chebyshev polynomials, such that higher order expansion coefficients are obtained through repeated matrix multiplications. It is this algorithm that is particularly suited for GPUs.

7.2.2 Retarded, advanced, principal, and non-principal Green's functions

The retarded Green's function can be written as [84]

$$G^R_{\mathbf{ij}}(t) = -i\theta(t)\sum_n \langle \Psi^{(n)}_{\mathbf{i}} | \Psi^{(n)}_{\mathbf{j}} \rangle^* e^{-iE_n t}. \tag{7.29}$$

Moreover, introducing a standard convergence factor $i\epsilon$ in the energy,[2] where ϵ is an infinitesimal positive number, and performing a Fourier transform, we find the Green's function in the energy domain

$$G^R_{\mathbf{ij}}(E) = -i\sum_n \langle \Psi^{(n)}_{\mathbf{i}} | \Psi^{(n)}_{\mathbf{j}} \rangle^* \int_0^\infty e^{i(E-E_n+i\epsilon)t} dt = \sum_n \frac{\langle \Psi^{(n)}_{\mathbf{i}} | \Psi^{(n)}_{\mathbf{j}} \rangle^*}{E - E_N + i\epsilon}. \tag{7.30}$$

[2] This can be motivated by the fact that the integral is carried out from 0 to ∞. If ϵ is small enough, it will only modulate the integrand far into the future, where it exponentially suppresses the integrand in order to make the integral converge. As physics close enough to $t = 0$ (around the time of interest) should not be affected by what is happening in the very far future, a modulation of the integrand that only is non-negligible in the very far future should not affect the result. In fact, if it did, our model would be very shaky to start with, because we would never be able to put a reasonable confidence into that our model actually corresponds to the experimental reality. For such confidence to exist in any model it has to allow for small but uncontrollable fluctuations locally in time, and large uncertainties in the very distant future and past. A corresponding convergence factor $-i\epsilon$ is similarly motivated for the advanced Green's function.

Taking the limit $\epsilon \to 0^+$ we have

$$G^R_{\mathbf{ij}}(E) = \sum_n \langle \Psi^{(n)}_{\mathbf{i}} | \Psi^{(n)}_{\mathbf{j}} \rangle^* \left(\frac{1}{E - E_n} - i\pi\delta(E - E_n) \right). \tag{7.31}$$

A similar derivation also leads to the advanced Green's function

$$G^A_{\mathbf{ij}}(E) = \sum_n \langle \Psi^{(n)}_{\mathbf{i}} | \Psi^{(n)}_{\mathbf{j}} \rangle^* \left(\frac{1}{E - E_n} + i\pi\delta(E - E_n) \right). \tag{7.32}$$

We can now define a "principal" and a "non-principal" Green's function as

$$\begin{aligned} G^P_{\mathbf{ij}}(E) =& \sum_n \langle \Psi^{(n)}_{\mathbf{i}} | \Psi^{(n)}_{\mathbf{j}} \rangle^* \frac{1}{E - E_n}, \\ G^{NP}_{\mathbf{ij}}(E) =& i\pi \sum_n \langle \Psi^{(n)}_{\mathbf{i}} | \Psi^{(n)}_{\mathbf{j}} \rangle^* \delta(E - E_n), \end{aligned} \tag{7.33}$$

such that the retarded and advanced Green's functions can be decomposed as

$$\begin{aligned} G^R_{\mathbf{ij}}(E) =& G^P_{\mathbf{ij}}(E) - G^{NP}_{\mathbf{ij}}(E), \\ G^A_{\mathbf{ij}}(E) =& G^P_{\mathbf{ij}}(E) + G^{NP}_{\mathbf{ij}}(E). \end{aligned} \tag{7.34}$$

In Section 7.2.1 we claimed that the "non-principal" Green's function is of primary interest. To understand why, we note that we intentionally started from the Green's function defined in the time domain, because in our point of view this is the Green's function that carries primary physical meaning. It is directly related to the propagation of particles from site **j** to **i** during the time interval t. The Green's function in the energy domain is in our view more of a mathematical convenience, and in particular we note that we had to introduce a convergence factor to actually arrive at it. Taking this view we should therefore be primarily interested in what happens to the retarded and advanced Green's functions as we perform the inverse Fourier transform to move them back into the time domain. In particular, we focus on the zero-time Green's function, as this is the Green's function of interest for most physical quantities. Because the inverse Fourier transform involves an integration from $-\infty$ to ∞, and the additional Fourier factor $e^{-iEt} = 1$ for $t = 0$, the contribution from the "principal" part drops out. This is why we say that the non-principal part is of primary interest.

To further substantiate this claim we note that two other commonly employed techniques have exactly the same outcome. First, integration around a contour, although in this case isolating the principal part through residue calculus, has the same effect as integrating the non-principal part along the segment of the real axis that is contained by the contour.[3] The principal part therefore

[3] This argument breaks down when the poles are located away from the real axis. However, we are only considering Hamiltonians that in principle can be diagonalized, giving rise to eigenstates with infinite life time.

serves an important mathematical purpose when contour techniques are employed, but it can be replaced by integration along the real energy axis when the principal and non-principal parts of the Green's function can be separated. Finally, we note that the commonly employed method of taking the imaginary part of the Green's function, such as when for example calculating the local density of states using

$$\rho(\mathbf{i}, E) = -\frac{1}{\pi}\mathrm{Im}\left(G_{\mathbf{ii}}(E)\right) \tag{7.35}$$

is yet another method for isolating the non-principal part. However, this is only guaranteed to work for diagonal entries of the Green's function as the factors $\langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^*$ can have imaginary components otherwise. Altogether this motivates why it in TBTK is possible to, in addition to calculate the retarded and advanced Green's functions, also calculate the "principal" and "non-principal" Green's functions. In particular, the "non-principal" Green's function is ideally suited for numerical integration.

7.2.3 Chebyshev expansion

For the moment we assume that the eigenstates of the system satisfy $E_n \in [-1, 1]$, and perform a Chebyshev transform of the non-principal Green's function

$$a_{\mathbf{ij}}^{(m)} = \int_{-1}^{1} G_{\mathbf{ij}}^{NP}(E)T_m(E)dE = i\pi\sum_n \langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^* T_m(E_n) = \sum_n a_{\mathbf{ij}}^{(nm)}, \tag{7.36}$$

where

$$a_{\mathbf{ij}}^{(nm)} = i\pi\int_{-1}^{1} \langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^* \delta(E - E_n)T_m(E)dE. \tag{7.37}$$

It follows that the Chebyshev expansion is given by

$$G_{\mathbf{ij}}^{NP}(E) = \frac{2}{\pi\sqrt{1-E^2}}\sum_{m=0}^{\infty}\frac{a_{\mathbf{ij}}^{(m)}}{1+\delta_{0m}}\cos(m\,\mathrm{acos}(E)), \tag{7.38}$$

which also can be written as

$$G_{\mathbf{ij}}^{NP}(E) = \sum_n g_{\mathbf{ij}}^{(n)NP} \tag{7.39}$$

where

$$g_{\mathbf{ij}}^{(n)NP}(E) = \frac{2}{\pi\sqrt{1-E^2}}\sum_{m=0}^{\infty}\frac{a_{\mathbf{ij}}^{(nm)}}{1+\delta_{0m}}\cos(m\,\mathrm{acos}(E)). \tag{7.40}$$

Noting that $a_{\mathbf{ij}}^{(nm)}$ is the Chebyshev transform of $i\pi\langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^*\delta(E-E_n)$, it follows that $g_{\mathbf{ij}}^{(n)NP}$ is the corresponding Chebyshev expansion. In particular, we can write

$$\begin{aligned} g_{\mathbf{ij}}^{(n)NP}(E) =& i\pi\langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^*\delta(E-E_n) \\ =& i\pi\langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^* \sum_m \frac{2a_{\mathbf{ij}}^{(nm)}}{i\pi^2\langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^*(1+\delta_{0m})} \frac{\cos(m\,\mathrm{acos}(E))}{\sqrt{1-E^2}}. \end{aligned} \tag{7.41}$$

Noting that the delta-function is real, we can now rewrite this factor as

$$\begin{aligned} \delta(E-E_n) =& \mathrm{Re}\left(\sum_{m=0}^{\infty} \frac{2a_{\mathbf{ij}}^{(nm)}}{i\pi^2\langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^*(1+\delta_{0m})} \frac{e^{\pm im\,\mathrm{acos}(E)}}{\sqrt{1-E^2}}\right) \\ =& \mathrm{Re}\left(\sum_{m=0}^{\infty} \frac{2a_{\mathbf{ij}}^{(nm)}}{i\pi^2\langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^*(1+\delta_{0m})} \frac{E^{\pm m}}{\sqrt{1-E^2}}\right), \end{aligned} \tag{7.42}$$

where the arbitrariness of the sign is a result of the real part of the expression inside the parenthesis being independent of the choice made. Written in this form the delta-function can be viewed as the real part of a holomorphic function resulting from analytic continuation in E. Further, knowing that the real part has a single delta-peak at E_n, it can be concluded that the expression also can be written as

$$\delta(E-E_n) = \mathrm{Re}\left(\frac{i}{E-E_n \pm i\epsilon}\right). \tag{7.43}$$

Having seen that the result of replacing $\cos(m\,\mathrm{acos}(E))$ by $e^{im\,\mathrm{acos}(E)}$ is to change

$$\delta(E-E_n) \to \frac{i}{E-E_n+i\epsilon}, \tag{7.44}$$

and thereby including the principal part, we understand that the principal Green's function can be obtained through the expansion

$$G_{\mathbf{ij}}^{P}(E) = \frac{2i}{\pi\sqrt{1-E^2}} \sum_{m=0}^{\infty} \frac{a_{\mathbf{ij}}^{(m)}}{1+\delta_{0m}} \sin(m\,\mathrm{acos}(E)). \tag{7.45}$$

We have thereby derived the Chebyshev expansion of both the "principal" and "non-principal" Green's functions, which through Eq. (7.34) also allows us to calculate the retarded and advanced Green's functions.

7.2.4 Matrix form

We have so far considered the Chebyshev transform of the Green's function when it is expressed as an ordinary function. We next consider the problem in matrix form. In particular, we consider expressions of the form

$$-i\theta(t)u^\dagger e^{-iHt}v, \tag{7.46}$$

where u and v are vectors with entries u_i and v_i, respectively, and H is the Hamiltonian. First of all, being Hermitian, H can be written on the form

$$UEU^{-1}, \tag{7.47}$$

where E is a diagonal matrix with diagonal entries $E_{jj} = E_j$, and U is unitary with entries U_{ij}. In particular, the ith column of U is the ith eigenvector $|\Psi^{(n)}\rangle$ of the Hamiltonian. The exponent can now be written as a power series

$$e^{-iHt} = \sum_m \frac{(-it)^m}{m!}H^m = U\left(\sum_{m=0}^{\infty}\frac{(-it)^m}{m!}E^m\right)U^{-1}, \tag{7.48}$$

where we have used that

$$H^m = (UEU^{-1})^m = UEU^{-1}(UEU^{-1})^{m-2}UEU^{-1} = UE^mU^{-1}. \tag{7.49}$$

Using this we can write

$$\begin{aligned} u^\dagger e^{-iHt}v &= \sum_j\left(\sum_i u_i^*U_{ij}\right)\left(\sum_{m=0}^{\infty}\frac{(-it)^m}{m!}E_j^m\right)\left(\sum_i U_{ij}^*v_i\right) \\ &= \sum_j\left(\sum_i u_i^*U_{ij}\right)e^{-iE_jt}\left(\sum_i U_{ij}^*v_i\right) \\ &= \sum_j \langle U_j(u)|U_j(v)\rangle e^{-iE_jt}, \end{aligned} \tag{7.50}$$

where we have used the notation $|U_j(v)\rangle = (U^{-1}v)_j$. In particular, if $v_{\mathbf{i}}$ is a vector with the only non-zero entry in row i, corresponding to the index $\mathbf{i}$, then $|U_n(v_{\mathbf{i}})\rangle = |\Psi_{\mathbf{i}}^{(n)*}\rangle$ and

$$-i\theta(t)u_{\mathbf{i}}^\dagger e^{-iHt}v_{\mathbf{j}} = -i\theta(t)\sum_n \langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^* e^{-iE_nt}. \tag{7.51}$$

We note that this is the same expression as Eq. (7.29). The Chebyshev transform of the corresponding "non-principal" Green's function can therefore be written as

$$a_{\mathbf{ij}}^{(m)} = i\pi\sum_n \langle\Psi_{\mathbf{i}}^{(n)}|\Psi_{\mathbf{j}}^{(n)}\rangle^* T_m(E_n) = i\pi u_{\mathbf{i}}^\dagger U T_m(E)U^{-1}v_{\mathbf{j}}, \tag{7.52}$$

where in the first equality we have used Eq. (7.36), while in the second equality we have used the definition of $u_{\mathbf{i}}$, $v_{\mathbf{i}}$, U, and that E is diagonal. Now, because $T_m(E)$ is a polynomial, and once more using that for each monomial of degree d we have $UE^dU^{-1} = H^d$, we can also write

$$a_{\mathbf{ij}}^{(m)} = i\pi u_{\mathbf{i}}^\dagger T_m(H) v_{\mathbf{j}}. \tag{7.53}$$

Having arrived at this form we can now define

$$|j_{\mathbf{i}}^{(0)}\rangle = v_{\mathbf{j}}, \tag{7.54}$$

and use the recursive properties of the Chebyshev polynomials to arrive at

$$a_{\mathbf{ij}}^{(m)} = i\pi \langle j_{\mathbf{i}}^{(0)} | T_m(H) | j_{\mathbf{j}}^{(0)} \rangle, \tag{7.55}$$

where

$$\begin{aligned} |j_{\mathbf{i}}^{(1)}\rangle &= H|j_{\mathbf{j}}^{(0)}\rangle, \\ |j_{\mathbf{i}}^{(m)}\rangle &= 2H|j_{\mathbf{j}}^{(m-1)}\rangle - |j_{\mathbf{j}}^{(m-2)}\rangle. \end{aligned} \tag{7.56}$$

These expressions define a recursive matrix-vector multiplication method for finding the Chebyshev expansion coefficients, and form the basis of the ChebyshevSolver.

7.3 Implementation details

For numerical reasons it is useful to reorganize the terms in the Chebyshev transform and expansion slightly, and we chose to define the Chebyshev transform as

$$b_{\mathbf{ij}} = \langle j_{\mathbf{i}}^{(0)} | T_m(H) | j_{\mathbf{j}}^{(0)} \rangle, \tag{7.57}$$

where

$$\begin{aligned} |j_{\mathbf{i}}^{(0)}\rangle &= v_{\mathbf{i}}, \\ |j_{\mathbf{i}}^{(1)}\rangle &= H|j_{\mathbf{j}}^{(0)}\rangle, \\ |j_{\mathbf{i}}^{(m)}\rangle &= 2H|j_{\mathbf{j}}^{(m-1)}\rangle - |j_{\mathbf{j}}^{(m-2)}\rangle. \end{aligned} \tag{7.58}$$

We also note that so far we have assumed that the energy spectrum is clamped to the interval $[-1, 1]$, because the Chebyshev transform is only defined on this interval. However, any Hamiltonian with a bounded spectrum can be rescaled as $\tilde{H} = H/s$, resulting in the same rescaling $\tilde{E} = E/s$ for the energies. Applying these modifications, it can be verified that the corresponding Chebyshev

expansion is

$$G_{\mathbf{ij}}^{NP}(E) = \sum_{m=0}^{\infty} \tilde{b}_{\mathbf{ij}}^{(m)} l_m(E),$$
$$l_m(E) = \frac{2i}{\sqrt{s^2 - E^2}(1 + \delta_{0m})} \cos(m \operatorname{acos}(E/s)), \tag{7.59}$$

where $\tilde{b}_{\mathbf{ij}}^{(m)}$ indicates that $\tilde{H}$ has been used to calculate the Chebyshev coefficients.

We are now ready to better understand some of the parameters that go into the creation of the ChebyshevSolver and CPropertyExtractor in Section 6.2. Namely,

```
SCALE_FACTOR
NUM_COEFFICIENTS
ENERGY_RESOLUTION
```

The scale factor is of course s, and should be chosen by the user to ensure that the energy spectrum is bounded by $[-1, 1]$. Further, the number of coefficients refers to how many expansion coefficients that should be used in the expansion, while the energy resolution determines the number of points at which the Green's function (and all derived quantities) is evaluated. We also mention that above we have collected all terms except the expansion coefficients into a quantity $l_m(E)$. This quantity is independent of the specific Green's function that is evaluated and can therefore be evaluated once, even if multiple Green's functions are calculated. This can significantly speed up the evaluation of the Green's functions because it reduces the number of mathematical functions that need to be evaluated. TBTK therefore allows these to be calculated and stored in a lookup table, and the third Boolean value (sixth parameter) passed to the CPropertyExtractor indicates whether this should be done or not. Finally, the two first Boolean values indicates whether the evaluation of Chebyshev expansion coefficients and the Green's function, respectively, should be performed on GPU(s).

8. Other developments

While TBTK is written to solve general problems, and the results described in the Results Part are focused at the specifics of topological superconductivity, we here describe some further method developments that are of intermediate generality. Because these also are described in the articles we are somewhat brief here, covering the general ideas rather than the details. In particular, we describe two developments related to the calculation of superconducting pair functions and general (potentially spin-polarized) currents.

8.1 Superconducting pair function

8.1.1 Classifying the pair function

In Section 4.7, we described how the superconducting pair function (order parameter) can be classified as s-, p-, d-wave, and so forth, based on how $F_{\mathbf{k}\sigma\sigma'}$ ($\Delta_{\mathbf{k}\sigma\sigma'}$) depends on the angular coordinate of $\mathbf{k}$. The classification is in analogy with the corresponding atomic orbitals, but has as is custom been described in momentum space, while the atomic orbitals have their angular dependence in real space. Here we show that as long as translational invariance is preserved, the classification can be done in real and momentum space interchangeably. However, once translational invariance is broken, the momentum space representation is less relevant, and the real space expression becomes of foremost importance. In particular, the real space expression forms the basis for our analysis of unconventional pair amplitudes in Paper III. Finally, we therefore describe how we in practice apply these ideas in lattice calculations, as well as how we extend this to what is known as odd-frequency pair amplitudes.

8.1.2 Spherical harmonics in real and momentum space

To make the connection between the real and momentum space classification clear, consider the expression

$$F_{\mathbf{R}\mathbf{r}\sigma\sigma'} = \langle c_{\mathbf{R}-\mathbf{r}\sigma} c_{\mathbf{R}+\mathbf{r}\sigma'} \rangle. \tag{8.1}$$

This is a completely general real space pair function in that the operators are free to sit at independent sites. Here we use $\mathbf{R}$ to denote a "center of mass" coordinate, while $\pm\mathbf{r}$ are vectors pointing from the "center of mass" to the

operator sites. In particular, we note that $F_{\mathbf{x}\sigma\sigma'} = F_{\mathbf{R}0\sigma\sigma'}$, with $\mathbf{R} = \mathbf{x}$, and where $F_{\mathbf{x}\sigma\sigma'}$ is the fully spin-dependent version of Eq. (4.42). The term "center of mass" coordinate is motivated by noting that $F_{\mathbf{R}\mathbf{r}\sigma\sigma'}$ contains information about the correlation between electrons at site $\mathbf{R}+\mathbf{r}$ and $\mathbf{R}-\mathbf{r}$. It can therefore be thought of as an electron pair wave function, with $\mathbf{R}$ being related to the propagation of the Cooper pairs center of mass, while $\mathbf{r}$ is related to the pairs motion relative to the center of mass. In particular, this motivates an expansion of the pair function of the form

$$F_{\mathbf{R}\mathbf{r}\sigma\sigma'} = \sum_{\mathbf{K}nlm} a^{\sigma\sigma'}_{nlm}(\mathbf{K},\mathbf{R}) j_n(|\mathbf{r}|) Y_{lm}(\theta_{\mathbf{r}},\varphi_{\mathbf{r}}) e^{i\mathbf{K}\cdot\mathbf{R}}, \tag{8.2}$$

where the $\mathbf{R}$ dependence has been expanded in the Fourier basis, and the $\mathbf{r}$ dependence have been expanded in the basis of spherical harmonics. Subscripts on the angular coordinates emphasises that these are the angular coordinates of $\mathbf{r}$. If we further restrict ourselves to the translationally invariant case, such that $F_{\mathbf{R}\mathbf{r}\sigma\sigma'} = F_{\mathbf{R}'\mathbf{r}\sigma\sigma'}$ for any $\mathbf{R}'$, implying $a^{\sigma\sigma'}_{nlm}(\mathbf{K},\mathbf{R}) = 0$ for every $\mathbf{K}$ except $\mathbf{K} = 0$, it follows that it is enough to consider

$$F_{0\mathbf{r}\sigma\sigma'} = \sum_{nlm} a^{\sigma\sigma'}_{nlm} j_n(|\mathbf{r}|) Y_{lm}(\theta_{\mathbf{r}},\varphi_{\mathbf{r}}), \tag{8.3}$$

where $a^{\sigma\sigma'}_{nlm} = a^{\sigma\sigma'}_{nlm}(0,0)$.

If we now instead perform the expansion for both variables in the Fourier basis, we have

$$F_{\mathbf{R}\mathbf{r}\sigma\sigma'} = \sum_{\mathbf{p}\mathbf{p}'} \langle c_{\mathbf{p}\sigma} c_{\mathbf{p}'\sigma'}\rangle e^{i\mathbf{p}\cdot(\mathbf{R}-\mathbf{r})} e^{i\mathbf{p}'\cdot(\mathbf{R}+\mathbf{r})} = \sum_{\mathbf{K}\mathbf{k}} \langle c_{\mathbf{K}-\mathbf{k}\sigma} c_{\mathbf{K}+\mathbf{k}\sigma'}\rangle e^{i2\mathbf{K}\cdot\mathbf{R}} e^{i2\mathbf{k}\cdot\mathbf{r}}, \tag{8.4}$$

where the last equality follows by defining $\mathbf{K} = (\mathbf{p}'+\mathbf{p})/2$ and $\mathbf{k} = (\mathbf{p}'-\mathbf{p})/2$. Once again assuming translational invariance, implying $\langle c_{\mathbf{K}-\mathbf{k}\sigma} c_{\mathbf{K}+\mathbf{k}\sigma'}\rangle = 0$ for all $\mathbf{K}$ except $\mathbf{K} = 0$, we have

$$F_{0\mathbf{r}\sigma\sigma'} = \sum_{\mathbf{k}} \langle c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'}\rangle e^{i2\mathbf{k}\cdot\mathbf{r}} = \sum_{\mathbf{k}nlm} b^{\sigma\sigma'}_{nlm} j_n(|\mathbf{k}|) Y_{lm}(\theta_{\mathbf{k}},\varphi_{\mathbf{k}}) e^{i2\mathbf{k}\cdot\mathbf{r}}, \tag{8.5}$$

where in the last equality we have expanded $\langle c_{-\mathbf{k}\sigma} c_{\mathbf{k}\sigma'}\rangle$ in spherical harmonics. If we further use the plane wave expansion [85]

$$e^{i2\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{lm} i^l j_l(2|\mathbf{k}||\mathbf{r}|) Y_{lm}(\theta_{\mathbf{k}},\varphi_{\mathbf{k}}) Y^*_{lm}(\theta_{\mathbf{r}},\varphi_{\mathbf{r}}), \tag{8.6}$$

the orthogonality condition for the spherical harmonics, and properly identify the expression with Eq. (8.3), we can write Eq. (8.5) as

$$F_{0\mathbf{r}\sigma\sigma'} = \sum_{nlm} a^{\sigma\sigma'}_{nlm} j_n(|\mathbf{r}|) Y_{lm}(\theta_{\mathbf{r}},\varphi_{\mathbf{r}}), \tag{8.7}$$

where $a^{\sigma\sigma'}_{nlm} j_n(|\mathbf{r}|) = 4\pi i^l b^{\sigma\sigma'}_{nlm} \sum_{|\mathbf{k}|} j_n(|\mathbf{k}|) j_l(2|\mathbf{k}||\mathbf{r}|)$.[1]

Now, Eq. (8.7) is of the same form as Eq. (8.3) and therefore provides a method for calculating the coefficients in the real space spherical harmonic basis from the corresponding momentum space basis. In particular, we note that only those spherical harmonics that occur in the momentum space basis occur in the real space basis. There is therefore a one-to-one correspondence between the spherical harmonics in real and momentum space as long as translational invariance is preserved. This means that as long as translational invariance is preserved, it is irrelevant whether the classification of the pair function occurs in real or momentum space.

8.1.3 Breaking translational invariance on a lattice

Once the translational invariance is broken, the momentum space representation loses much of its power. It is therefore more useful to consider the real space expression directly. In particular, in Paper III we consider the pair function in a two-dimensional topological superconductor with a vortex, where the presence of a vortex and edges breaks the translational invariance. Moreover, because we perform a lattice calculation in two dimensions, while the spherical harmonics are basis functions for a continuous three-dimensional space, we need to find a corresponding discrete two-dimensional analog for performing the classification.

In Paper III we choose basis functions such that s-wave refers to the case $\mathbf{r} = 0$, in agreement with the terminology in Section 4.6.1. We also define an extended s-wave basis function that is a non-zero constant on the nearest neighbor sites of $\mathbf{R}$. Further, for p-wave we choose basis functions that are non-zero on the nearest neighbor sites, and that have the characteristic p-wave dependence $e^{\pm i\varphi_\mathbf{r}}$. Finally, for d-wave we choose basis functions that are non-zero on the nearest neighbor and next nearest neighbor sites, and that varies as $e^{\pm i2\varphi_\mathbf{r}}$. These choices correspond to the most localized basis functions that display full s-, p-, and d-wave characteristics. We note that the choices involves a simultaneous choice of angular and radial bahavior, which are separated into $Y_{lm}(\theta_\mathbf{r}, \varphi_\mathbf{r})$ and $j_n(|\mathbf{r}|)$, respectively, in the three-dimensional continuum expression. Having chosen basis functions, we then project the pair function onto these in order to extract relevant pair amplitudes at different sites in the lattice. For more details on these calculations we refer the reader to the section "VII. Pair amplitudes" in Paper III.

[1] We have in this section used the summation symbol to interchangeably denote summation and integration, depending on if the variables are continuous or discrete. The symbol $\sum_{|\mathbf{k}|}$ should therefore be understood to be the radial integral, or the corresponding discrete sum.

8.1.4 Odd-frequency pairing

In Section 4.7 we described how pair functions that are even under intechange of momentum, such as s- and d-wave, has to be odd in spin (singlet), while pair functions that are odd in momentum has to be even in spin (triplet). However, this can be violated if the behavior of the pair function in time is taken into account [86–91]. The pair functions considered so far have all been with operators at equal time, and in particular this means that we have been looking at the components of the pair functions that are even in time, as odd functions are zero at the point they are odd around. If also the time dependence is taken into account, the pair function has to be totally odd under the simultaneous interchange of momentum, spin, and time. In Paper III we also present results for odd-frequency pairing by taking the time derivative of the pair functions described above, because odd functions have a non-zero derivative at the point they are odd around.

8.2 General currents

8.2.1 When spin is not a good quantum number

In Paper IV and V we investigate currents induced around magnetic impurities at the surface of an s-wave superconductor with Rashba spin-orbit interaction. Because the combination of magnetic impurities and Rashba spin-orbit interaction gives rise to terms in the Hamiltonian involving all spin axes, spin is not a good quantum number. Moreover, we are interested in spin-polarized currents that can be polarized along one of several different axes. For this reason a general method for calculating spin-polarized currents had to be developed and implemented. Here we describe the main ideas, and refer to Appendix B in Paper V for details and proofs.

8.2.2 The Heisenberg equation

The calculation of spin-polarized currents is based on the Heisenberg equation of motion

$$\frac{d\hat{\rho}_{\mathbf{x}\sigma}}{dt} = \frac{i}{\hbar}\left[H, \rho_{\mathbf{x}\sigma}\right], \tag{8.8}$$

where $\hat{\rho}_{\mathbf{x}\sigma} = c^\dagger_{\mathbf{x}\sigma} c_{\mathbf{x}\sigma}$ is the density operator for σ-spins at site $\mathbf{x}$. In principle, the right hand side can be expanded and terms responsible for moving σ-spins in and out of site $\mathbf{x}$ can be identified as responsible for generating currents. For example, terms of the form $c^\dagger_{\mathbf{x}\sigma} c_{\mathbf{x}+\mathbf{b}\sigma}$ and $c^\dagger_{\mathbf{x}+\mathbf{b}\sigma} c_{\mathbf{x}\sigma}$ move spins in and out of the site $\mathbf{x}$, respectively, along the bond pointing in the direction $\mathbf{b}$, because the operator pairs remove electrons from the site corresponding to the annihilation operator and recreates them at the site corresponding to the creation operator.

The current is then evaluated by taking expectation values of these operators. This is indeed the basic idea behind our derivation. However, straightforward evaluation of the commutator produces a wealth of terms when the Hamiltonian is complex enough. Moreover, we are interested in arbitrary directions of the spin σ, which in practice means that the operators with sigma indices are not necessarily written in the same spin-basis as the Hamiltonian. A general treatment therefore requires the problem to be addressed with a certain amount of systematic book keeping.

Without going into the details of the derivation, we now highlight the main results and notation that is introduced through our derivation. Moreover, we will see that this results in a natural division of the problem into a two stage process, where relevant quantities first are calculated in the basis of the Hamiltonian, which then are used in the second stage to calculate currents in an arbitrary spin-basis. While the first stage is computationally demanding, it is done without fixing the spin-polarization axis of interest. The second stage can then quickly be run multiple times for different spin-polarization axes. This can significantly reduce the time needed to get a complete understanding of the system.

8.2.3 Sink-source, spin current, and spin-flipping current

When the commutator in Eq. (8.8) is expanded, the resulting expression can be written as

$$\frac{d\hat{\rho}_{\mathbf{x}\sigma}}{dt} = \hat{S}_{\mathbf{x}\sigma} - \sum_{\mathbf{b}\neq 0} \left(\hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\sigma} + \hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\bar{\sigma}} \right). \tag{8.9}$$

Here $\hat{S}_{\mathbf{x}\sigma}$ is an on-site sink-source term that arise as a result of local spin-flips. Further, $\hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\sigma}$ carries electrons with spin σ from site $\mathbf{x}$ along bond $\mathbf{b}$, and is therefore an ordinary spin-current. Finally, $\hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\bar{\sigma}}$ similarly carries electrons with spin σ from site $\mathbf{x}$ along bond $\mathbf{b}$, but flips the spin to the oposite spin $\bar{\sigma}$ in the process. For the purpose of currents, it is the last two terms that are of interest, but for completeness we treat all three terms.

The actual values of the sink-source term and currents are calculated as the expectation values of the corresponding operators. In Appendix B of Paper V

we show that these can be written as

$$\langle \hat{S}_{\mathbf{x}\sigma} \rangle = 2\mathrm{Re}\left(\sum_{\kappa\kappa'} a^{\bar{\sigma}\sigma\sigma\sigma}_{\mathbf{x},\bar{\sigma},\mathbf{x},\sigma} \bar{b}^*_\kappa b_{\kappa'} \langle c^\dagger_{\mathbf{x}\kappa} c_{\mathbf{x}\kappa'} \rangle_H \right), \tag{8.10}$$

$$\langle \hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\sigma} \rangle = -2\mathrm{Re}\left(a^{\sigma\sigma\sigma\sigma}_{\mathbf{x}+\mathbf{b},\sigma,\mathbf{x},\sigma} \sum_{\kappa\kappa'} b^*_\kappa b_{\kappa'} \langle c^\dagger_{\mathbf{x}+\mathbf{b},\kappa} c_{\mathbf{x}\kappa'} \rangle_H \right), \tag{8.11}$$

$$\langle \hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\bar{\sigma}} \rangle = -2\mathrm{Re}\left(a^{\bar{\sigma}\sigma\sigma\sigma}_{\mathbf{x}+\mathbf{b},\bar{\sigma},\mathbf{x},\sigma} \sum_{\kappa\kappa'} \bar{b}^*_\kappa b_{\kappa'} \langle c^\dagger_{\mathbf{x}+\mathbf{b},\kappa} c_{\mathbf{x}\kappa'} \rangle_H \right). \tag{8.12}$$

Here the expectation values on the left side refers to the final expressions in the basis of choice, while the expectation values with subscript H refers to the expectation values in the basis of the Hamiltonian. Further, the a's are coefficients that in turn are calculated as sums and products of the b's above, and another set of A coefficients that are evaluated from the Hamiltonian. We do not go into details here, but instead refer the interested reader to Appendix B in Paper V. However, the b coefficients are related to the choice of spin-polarization axis, and are given by

$$b_\uparrow = \cos\left(\frac{\theta}{2}\right), \tag{8.13}$$

$$b_\downarrow = \sin\left(\frac{\theta}{2}\right) e^{-i\varphi}, \tag{8.14}$$

where θ and φ are the spherical coordinates of the spin polarization axis relative to the coordinate system of the Hamiltonian, and $\uparrow$ ($\downarrow$) refers to the spin that is up (down) with respect to this axis.

8.2.4 Vector current

The two quantities $\langle \hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\sigma} \rangle$ and $\langle \hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\bar{\sigma}} \rangle$ are the spin-polarized and spin-flipping currents, respectively, along the bond from $\mathbf{x}$ to $\mathbf{x}+\mathbf{b}$. These are expressed as scalar quantities defined on the bonds between sites, but in practice we are interested in the currents as a vector field defined on the sites. For this reason we also define the current at site $\mathbf{x}$ as the directed average of the currents along the bonds connected to site $\mathbf{x}$ as

$$\mathbf{J}_{\mathbf{x}\sigma\sigma'} = \frac{1}{2} \sum_{\mathbf{b}\neq 0} \mathbf{b} \langle \hat{J}^{\mathbf{b}}_{\mathbf{x}\sigma\sigma'} \rangle, \tag{8.15}$$

where $\sigma' \in \{\sigma, \bar{\sigma}\}$.

8.2.5 Two stage calculation

As mentioned earlier, the evaluation of the currents is divided into a two stage process that allows for more efficient numerical evaluation. We can now un-

derstand how by noting that the expressions in Section 8.2.3 involves three types of coefficients a, A, and b, and two expectation values $\langle ... \rangle$ and $\langle ... \rangle_H$. We mentioned that the coefficients A and expectation values $\langle ... \rangle_H$ are calculated from the Hamiltonian (or in the basis of the Hamiltonian), and these are therefore calculated in stage one. In particular, the evaluation of the expectation values $\langle ... \rangle_H$ is a computationally demanding problem. In stage two, a specific choice is made for the spin axis by specifying the b coefficients, from which the a coefficients, the expectation values $\langle ... \rangle$, and the vector currents then follows by comparative computational ease.

Part IV:
Results

9. Results

As described in Section 5.2, recent theoretical and experimental progress have indicated that topological superconductivity and accompanying Majorana fermions can be engineered through the use of material building blocks exhibiting s-wave superconductivity, Rashba spin-orbit interaction, and magnetism. Further, in Section 5.4.2 we introduced a model Hamiltonian that describes such a set-up. The exact experimental realization of the model is not specified, and can therefore be thought of as equally well describing one of several different realizations. For example: magnetic impurities deposited at the surface of an s-wave superconductor with Rashba spin-orbit interaction coming from surface effects, a semiconductor in contact with a superconductor and a ferromagnet, or a similar semiconductor-superconductor heterostructure with the ferromagnetic layer replaced by an external magnetic field. The results that we present in Paper I-VII are all in one way or another based on slight variations of this Hamiltonian. Here we describe these results.

9.1 Model Hamiltonian

While the model Hamiltonian essentially was introduced in Section 5.4.2, we here write it in its full generality

$$\mathcal{H} = \mathcal{H}_{kin} + \mathcal{H}_{V_Z} + \mathcal{H}_{SO} + \mathcal{H}_{SC}, \tag{9.1}$$

$$\mathcal{H}_{kin} = -t \sum_{\langle \mathbf{i},\mathbf{j} \rangle \sigma} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{j}\sigma} - \mu \sum_{\mathbf{i}\sigma} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{i}\sigma}, \tag{9.2}$$

$$\mathcal{H}_{V_Z} = -\sum_{\mathbf{i}\sigma\sigma'} \left[V_Z(\mathbf{i}) \hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma} \right]_{\sigma\sigma'} c^\dagger_{\mathbf{i}\sigma} c_{\mathbf{i}\sigma'}, \tag{9.3}$$

$$\mathcal{H}_{SO} = \alpha \sum_{\mathbf{ib}} \left(e^{i\theta_{\mathbf{b}}} c^\dagger_{\mathbf{i}+\mathbf{b}\downarrow} c_{\mathbf{i}\uparrow} + \text{H.c.} \right), \tag{9.4}$$

$$\mathcal{H}_{SC} = \sum_{\mathbf{i}} \left(\Delta(\mathbf{i}) c^\dagger_{\mathbf{i}\uparrow} c^\dagger_{\mathbf{i}\downarrow} + \text{H.c.} \right), \tag{9.5}$$

where t is the hopping amplitude, μ is the chemical potential, V_Z the Zeeman term, α the spin-orbit interaction, $\mathbf{b}$ a nearest neighbor bond along the direction $\theta_{\mathbf{b}}$, and $\Delta_{\mathbf{i}}$ the superconducting order parameter. The additional changes to the Hamiltonian presented in Section 5.4.2 is that the Zeeman term has been given spatial dependence, and that it now can point in any given direction $\hat{\mathbf{n}}$.

In Paper I-III we consider the case of a homogeneous Zeeman term pointing perpendicular to the two-dimensional plane of the lattice. In Paper IV and V, we consider the case of magnetic impurities in point, line, and block configurations. In this case $V_Z(\mathbf{i})$ is set to zero everywhere except at the impurity sites, where it is set to a non-zero constant value. In these papers also the role of the direction $\hat{\mathbf{n}}$ of the Zeeman term is investigated. Spatial variations of the Zeeman term are also considered in Paper VI and VII.

The papers also differ in the specific order parameter configuration that is chosen. In Paper I and III a vortex is studied, which, in line with the discussion in Section 4.6.3, is modelled by twisting the phase of the order parameter such that $\Delta(\mathbf{i}) = |\Delta(\mathbf{i})|e^{i\theta_\mathbf{i}}$, where $\theta_\mathbf{i}$ is the polar coordinate of $\mathbf{i}$. In the other papers no such twist is considered. The order parameter is also in most papers, in line with the discussion in Section 4.5, calculated self-consistently using

$$\Delta(\mathbf{i}) = -V_{SC} \sum_{E_\nu<0} v_{\mathbf{i}\downarrow}^{(\nu)*} u_{\mathbf{i}\uparrow}^{(\nu)}, \tag{9.6}$$

where V_{SC} is the superconducting pair potential in the s-wave singlet channel.

9.2 Main objectives and limitations

Having discussed the model that unifies the results presented in Paper I-VII, we now turn to the main objectives and limitations of these papers. Many of our basic results, in particular the existence of Majorana fermions in vortex cores and at wire endpoints, have already been predicted in analytical continuum calculations using simpler models. A reoccurring theme is therefore to numerically investigate these phenomena using a more realistic model, which in particular means solving self-consistently for the order parameter on a lattice. Analytical models are useful for capturing and explaining essential phenomena, but because they often need to be particularly simple to be possible to solve, it is an important question whether they actually describe a relevant experimental reality.

We note that while we do aim at coming closer to the experimental reality, we do not claim to faithfully describe any particular experimental setup. In our view the two largest discrepancies between an actual system and the models used is likely to be the lack of a realistic band structure and the usage of large model parameters. In particular, using a single band model that results in an energy dispersion of the form $E = -2t(\cos(k_x) + \cos(k_y))$ and the unrealistically large size of the order parameter $|\Delta| > 0.1t$. The first of these discrepancies has its origin in keeping the model general, not introducing material specific parameters. The second is due to numerical limitations, where more realistic parameter values would require the calculations to be performed

on larger models.[1] We therefore do not claim to parameter wise be close to the experimental reality, but rather complexity wise: self-consistently relax the order parameter to obtain a realistic profile, the presence of a lattice, multiple impurities without a particular symmetric arrangement, lower dimensional topological superconductor structures such as one-dimensional wires embedded in a two-dimensional superconductor, and so forth. Self-consistency also allows us to phenomenologically study the effect of other model parameters such as the Zeeman term and Rashba spin-orbit interaction on the order parameter. Finally, performing the calculation for multiple parameter values allows for the robustness of the phenomena to be probed, hinting at their validity also for more realistic parameter values.

While verifying and testing the limits of results obtained in simpler models is one objective, it is far from the only one. For example, in Paper I we map out a phase diagram that provides important information about the effect of the Zeeman term and Rashba spin-orbit interaction on the stability of the topologically non-trivial phase. The phase diagram also provides values that allow for subsequent numerical calculations to be performed in a relevant parameter region. In Paper II, we relate the more abstract classification of the topological phase using a Chern number to a spin Skyrmion structure in the band structure. This can be used to directly measure the topological phase using spin-polarized ARPES. Possible experimental signatures of Majorana fermions in vortex cores, such as local density of states measurements using scanning tunneling microscopy and their relation to similar signatures in the spectral function are investigated in Paper III. Here also the presence of unconventional and odd-frequency pair functions in vortices with and without Majorana fermions, and in the surrounding bulk is investigated. Paper IV and V are concerned with (spin-polarized) currents induced around magnetic impurities in point, line, and block config-

[1] This is due to the number of eigenstates being the same as the size of the Hilbert space. A larger lattice size leads to a larger Hilbert space, and therefore more eigenstates. Because the energy spectrum typically is bounded to an interval the size of the band width, the energy resolution roughly goes as the band width divided by the number of eigenstates. For a reasonable amount of trust to be put into the results, the calculated quantities therefore should be measured on an energy scale that is significantly larger than this resolution. Particularly so for collective quantities such as the order parameter, which arise from the collective behavior of many eigenstates. A typical size of the Hilbert space for results reported in the Papers that are based on diagonalization (I-V and VII) is 6400 and takes $2-3$ days to generate self-consistently, while a typical size used during the iterative research phase is 1600, taking one to a few hours to solve. With a band width of $8t$, this means a smallest energy resolution of $0.00125t$ and $0.005t$, respectively, and that the energy scale of the order parameter $|\Delta| > 0.1t$ is at least one to two orders of magnitude larger than the order of the numerical errors. For diagonalization the execution time cannot be improved significantly because self-consistent calculations using diagonalization is a problem with few opportunities for effective parallelization, and real world time and memory constraints therefore sets a limit to the numerical accuracy. In fact, this is one of the reasons we have shown interest in also investigating the possibilities of using the Chebyshev expansion of the Green's function, as it allows for notably larger model sizes to be considered. However, this is a recent development, and except for the results presented in Paper VI, we rely on diagonalization.

urations at the surface of s-wave superconductors with Rashba spin-orbit interaction. These results have a broader focus than being restricted to the topic of topological superconductivity and Majorana fermions. However, they serve an important role in understanding the claim that Majorana fermions at the end points of one-dimensional wires are spin-polarized. Further, in Paper VI we investigate a network of one-dimensional wires embedded in a two-dimensional superconductor, showing that Majorana fermions not only appear at wire end points, but also at junctions between an odd number of wire segments. This is particularly done with the intention of providing an experimental way of strengthening (or weakening) the argument that zero energy bias peakes experimentally found and interpreted as being due to Majorana fermions indeed are so. Finally, in Paper VII we investigate a phenomenon not directly related to topological superconductivity, where magnetic impurities can give rise to a local π-phase shift in the order parameter, with the objective to provide a theoretical understanding of its origin. While this summarizes the main objetives of our work, we next go into some more detail about the main results of these papers.

9.3 Main results

9.3.1 Paper I

We confirm earlier predictions [25, 29, 30, 77] that Majorana fermions are supported in vortex cores in the topologically non-trivial phase. We also numerically map out a phase diagram, addressing questions related to the size of the topologically non-trivial phase as a function of the pair potential, Zeeman term, and Rashba spin-orbit interaction. In particular, we find that while the Zeeman term naturally is detrimental to superconductivity, the Rashba spin-orbit interaction helps their coexistence. This is very important, as the topologically non-trivial phase otherwise would not exist. While the Rashba spin-orbit interaction originally is included in the model for topological reasons, as described in Chapter 3 and 5, this is an additional important role it plays that is not seen in non self-consistent calculations.

In our phase diagram we identify three regions that we label I, II, and III, that corresponds to the topologically trivial superconducting phase, topologically non-trivial superconducting phase, and the non-superconducting phase. In addition, we find a fourth region that we label I' that is topologically trivial, but for which vortex cores become ferromagnetic and contain modes with energy close to zero that are not isolated Majorana fermions. This is important from an experimental point of view because it explicitly demonstrates that false positive signatures of isolated Majorana fermions is a very real possibility.

Finally, we also find spectral asymmetries in the model with respect to the Zeeman term and vortex rotation direction. These are traced back to the interplay between the Zeeman term, Rashba spin-orbit interaction, and vortex

rotation direction. While the rotation direction sets a direction of motion in the system, this motion couples to the spin through the Rashba spin-orbit interaction, and the spin is in turn coupled to the Zeeman term. The result is that the spectrum becomes a function of both the rotation direction and Zeeman term, and is symmetric only under a simultaneous reversal of both.

9.3.2 Paper II

As mentioned in Section 5.4.3, the topological phases of the model have been classified using a Chern number [30]. Further, from Section 2.4.4 we know that the geometrical meaning of the Chern number is related to the twist of a complex valued fiber bundle. This is admittedly rather abstract, but when the fiber has two complex dimensions, it is possible to give the Chern number a more geometrically intuitive interpretation. It is well known that the Chern number, through an appropriate geometrical mapping, also can be interpreted as a Skyrmion number [92]. However, in the present model the Chern number is used to classify a fiber bundle associated with a four-by-four matrix, which implies that the fiber has four complex dimensions and it is therefore not immediately clear that a similar construction is possible.

In this paper we show that also for this model it is possible to arrive at a Skyrmion number classification. In particular, this Skyrmion manifests itself in the form of a twisted spin structure associated with each band in momentum space. While the trivial phase have a trivial spin structure in each band, the non-trivial phase is distinguished by the collection of spins from a single band close to the Fermi level forming a hedgehog. This provides a potential way to measure the topological invariant directly using, for example, spin-polarized ARPES [93]. We also show that the Skyrmion spin structure is beneficial because it can be generalized to the case of a one-dimensional wire. In this case the spins are restricted to a great circle, where the non-trivial and trivial phases are distinguished by whether the collection of all spins twist around this circle or not. From a conceptual point of view this is very helpful, because it makes the connection between the physics of the one- and two-dimensional cases more explicit, an objective that we in a different way also pursue in Paper V.

9.3.3 Paper III

In Paper III we calculate the local density of states along a line cutting through a vortex core, and relate the results to signatures of the topological phase in the spectral function. These results are particularly focused at providing guidance for experimental investigations of Majorana fermions in vortex cores. The local density of states can be measured with for example STM/STS, while the spectral function can be measured with ARPES. We perform a careful analysis

of the bulk band structure, relating signatures in the the local density of states and spectral function to the shape of the band structure. In particular, the non-trivial phase have a Mexican hat shaped bulk band structure, giving rise to two band edges that can be seen directly in the spectral function, and indirectly gives rise to signatures also in the local density of states. Most notably, the Mexican hat shaped band structure can lead to an apparent increase in the band gap in the vortex core. This is due to that the lower band edge locally collapses in the vortex core to give rise to Majorana fermions and so called Caroli-de Gennes-Matricon states [94], while the upper band edge which exists also in the absence of superconductivity survives. The superconducting gap therefore collapses as expected, but a straight forward investigation of the local density of states can give the opposite impression. This and related signatures can potentially be used to verify that a system is in the non-trivial phase.

Due to claims that Majorana fermions are related to odd-frequency pairing [86–89], and a general interest in unconventional pairing in vortices [95–99], we also calculate such pair amplitudes based on the principles outlined in Section 8.1. We find a wealth of unconventional and odd-frequency pair amplitudes, and in particular analyze how the question of whether they occur or not can be understood from the total angular momentum of their orbital and spin component. However, while there is no shortage of odd-frequency pair amplitudes, we do not find a strong correlation with the onset of the topologically non-trivial phase and the appearance of Majorana fermions.

9.3.4 Paper IV and V

Both Paper IV and V are concerned with currents around magnetic impurities on the surface of s-wave superconductors. In Paper IV we show that a single magnetic impurity gives rise to localized spin-polarized currents in its vicinity, given that the Rashba spin-orbit interaction is non-zero. This can be understood as a consequence of the magnetic impurity coupling to the spin of the electrons, while the Rashba spin-orbit interaction in turn couples the spin to the momentum and therefore sets up a preferred rotation direction. We note that this have some similarities with the interplay between the vortex rotation direction and Zeeman term in Paper I. The focus of Paper IV is to understand the individual impurities that act as building blocks of topological superconductors constructed with the help of such impurities. In particular, this means extending the understanding of magnetic impurities in s-wave superconductors, building on top of a since long well established understanding of such impurities in terms of so called Yu-Shiba-Rusinov (YSR) states [100–102].

In paper V a more thorough investigation of the same phenomenon is carried out for line and block configurations of magnetic impurities. While Paper IV is aimed at understanding the individual building blocks of a topological superconductor, Paper V instead is concerned with the one- and two-dimensional

systems themselves. Spin-polarized currents are shown to flow around the two dimensional block, and their relation to the topologically protected edge states are described. It is also shown through local density of states calculations that the Majorana fermions that appear at the end points of one-dimensional wires are spin-polarized. We explain this in relation to the spin-polarized currents of the two-dimensional block. In this sense Paper V can be seen as taking the opposite approach of Paper IV, which focuses on the zero-dimensional building blocks, explaining the physics of the one-dimensional wire through dimensional reduction of the topological band theory of the two-dimensional block.

9.3.5 Paper VI

We show that Majorana fermions appear not only at end points of one-dimensional wires, but also at junctions between an odd number of wire segments, while no Majorana fermions appear at junctions between an even number of segments. The effect can be understood as a consequence of the fact that Majorana fermions have to come in pairs, and therefore only can annihilate in pairs. When an odd number of end points is brought together to form a junction, only an even number of Majorana fermions can be hybridized, leaving an odd number of Majorana fermions at such junctions. Experimentally this is a valuable signature that can be probed with for example STM. If the even-odd effect can be seen, the case for that the zero energy bias peaks are Majorana fermions is much stronger than current results that have found such peaks only at end points.

To provide further guidance for experiments, we also show how the results depend on the model parameters. Most notably, we show how the excitation gap varies with the Rashba spin-orbit interaction and Zeeman term, providing further signatures for identifying the topologically non-trivial phase. In addition, this also tells us how the excitation gap varies with the model parameters. As mentioned in Section 5.5.2, a large excitation gap is important if the Majorana fermions are to be used for topological quantum computation. We also mention that Paper VI is the only paper in which the Chebyshev method have been used. This is notably manifested in the system size of 400×400 lattice sites that is used in this study, which can be compared to the typical lattice size of up to 40×40 in other papers.

9.3.6 Paper VII

This is the paper that has least direct connection to topological superconductivity. It addresses an old issue related to the YSR states discussed in Paper IV, where previous studies have found that the local order parameter can acquire a local π-shift at a magnetic impurity [103–106]. We consider a line of

such impurities and show that the phase shift can be traced back to a certain type of resonance between the Bogoliubov-de Gennes quasi-particles. While all occupied states tend to be in phase with each other and the condensate, the unoccupied states are out of phase. The presence of a magnetic impurity pulls some of the originally unoccupied states into the condensate, simultaneously expelling some originally occupied states, locally giving rise to a π-phase shift. The problem is first considered without a Rashba spin-orbit interaction to obtain a basic understanding of the phenomenon. However, motivated by the presence of Rashba spin-orbit interaction in topological superconductors, we also show how the results are modified when it is included.

10. Summary and outlook

Topological phases of matter is a fascinating and rapidly expanding field of condensed matter physics. As a theoretical framework it is interesting because of the many formal analogies that can be drawn between theories for high energy physics and condensed matter. This means that many theoretical concepts, previously developed for understanding physics on a more fundamental level, now also can be used to understand the quasi-particle excitations that occur in condensed matter physics. Topological phases of matter therefore provide an excellent opportunity to leverage decades of research into high energy physics, to make predictions about physics on the nanometer length scale and above. In particular, it has the potential to provide immediate impact to society, through the manufacturing of devices that take advantage of material properties that are protected by topology.

In the work leading up to this thesis, the focus has been on topological superconductors that can be manufactured by combining materials exhibiting superconductivity, magnetism, and Rashba spin-orbit interaction. These have potential application in the form of topological quantum computation devices. One of the core objectives of the papers included in this thesis is to investigate realistic lattice models of topological superconductors, solving self-consistently for the order parameter. This allows for more complex models to be considered than more idealized analytical ones. In particular, we address issues related to the stability of the topologically non-trivial phase, existence of Majorana fermions, measurability of topological invariants, signatures of Majorana fermions in local density of states and spectral function measurements, and more. In addition, we also investigate phenomena such as the appearance of spin-polarized currents and a π-phase shift in the order parameter around magnetic impurities. These later results are of broader interest, not strictly related to topology alone, but is of particular interest for understanding the larger context within which topological superconductivity exists.

While the results are aimed at topological superconductivity and Majorana fermions, the work have been performed in parallel with the evolution of a numerical library for solving general bilinear Hamiltonians. This has resulted in a c++ library currently consisting of more than 15,000 lines of code, mostly written in c++, which is freely available online, and can be used to quickly setup and solve models of arbitrary complexity. In particular, the library allows for both wave function and single particle Green's function methods to be employed for calculating quantities. Several future directions are therefore possible. First, the library can be built upon by leveraging analytical methods

in combination with the possibility of calculating single particle Green's functions on demand, to extend the range of problems that are possible to solve to include also non-bilinear Hamiltonians. This would allow for many-body interactions to be taken into account, be it in topological superconductors, or other systems. A second possible direction is to utilize the ability to easily setup models of arbitrary complexity to study more complicated lattice structures or geometries, including systems consisting of multiple subsystem of different dimension, such as a wire that is placed on top of a superconductor rather than being embedded inside it. However, the generality does not stop at topological superconductors. The methods have in particular been constructed with the intention of also allowing the study of molecules of arbitrary complexity, molecules in connection with substrates, or lattices with arbitrary shapes, to name a few.

11. Topologisk bandteori och Majoranafermioner (Summary in Swedish)

Det har sedan tidigt 1700-tal varit känt att vissa material leder elektrisk ström väl, medan andra är mycket dåliga ledare. Många av de material som har god ledningsförmåga är metaller, i vilka atomerna är regelbundet organiserade i en periodisk kristallstruktur. Samtidigt är många andra kristalina material mycket goda isolatorer, trots att de i sin struktur liknar metallerna. Det dröjde ända till 1920-talets slut innan den då nyupptäckta kvantmekaniken till slut kunde tillämpas för att förklara vad som ger upphov till så drastiskt olika elektriska egenskaper hos dessa i övrigt mycket liknande typer av material. Den resulterande teorin kallas för bandteori, och även om inte alla ledare och isolatorers egenskaper kan förklaras med hjälp av denna teori, så är bandteori ett av de mest centrala koncepten inom dagens materialfysik.

Det viktigaste begreppet inom bandteori är materialens bandstruktur. Denna innehåller bland annat information om hur mycket energi som behöver tillföras elektronerna inuti materialet innan dessa kan bidra till en ström. Om bandstrukturen visar att det är möjligt att sätta fart på elektronerna genom att tillföra godtyckligt små mängder energi, då är materialet en metall. Det är dock möjligt att det finns en nedre gräns för hur mycket energi som behöver tillföras en elektron för att få denna att bidra till en ström. Om detta är fallet så säger man att ett bandgap separerar de så kallade valenstillstånd elektronerna normalt är i (valensband), från de tillstånd som kan bidra till en ström (ledningsband). Om detta bandgap är stort blir ledningsförmågan mycket dålig och materialet är en isolator, men om det istället är litet men inte noll så är det en så kallad halvledare.

Ett tillstånd som i många material uppstår vid låga temperaturer är supraledning. Likt det metalliska tillståndet är supraledare goda ledare, men de har förvåndande nog också stora likheter med isolatorer. På grund av ofrånkomliga orenheter och ljudvågor i metaller upplever dessa motstånd som gör att strömmar inuti materialet avtar, om inte strömen kontinuerligt drivs på av en yttre spänning. I motsats till detta är supraledare perfekta ledare som kan bevara strömmar som sats igång inuti dem. Detta kan i likhet med isolatorernas egenskaper förklaras med en typ av bandgap. I en isolator hindrar bandgapet elektronerna från att bli ledningselektroner. För en supraledare är det istället så att så snart en ström satts igång, så hindrar supraledarens bandgap elektronerna från att sluta bidra till denna ström.

Trots att både bandteori och den teorin som förklarar dessa supraledande effekter varit kända sedan ca 80 och 50 år tillbaka i tiden, så har det nyligen

upptäckts att bandteorier kan ha så kallade topologiska egenskaper. För vissa typer av isolatorer innebär detta att trots att materialet egentligen är en isolator i sitt inre, så leder det strömmar kring sitt yttre. Dessa strömmar har vidare en väldigt specifik ledningsförmåga, vilken är okänslig för många typer av orenheter som kan förekomma i materialet.

För supraledare är istället den mest tilltalande konsekvensen av de topologiska egenskaperna att dessa förutsäger att vissa typer av supraledare ger upphov till så kallade Majoranafermioner. En Majoranafermion är en i partikelfysik hypotetisk typ av fundamentalpartikel. Det är än så länge oklart om det verkligen finns fundamentalpartiklar som är Majoranafermioner eller inte, men i supraledare med de rätta topologiska egenskaperna skulle dessa kunna uppstå som så kallade kvasipartiklar. Det vill säga, inte som verkligt fundamentala partiklar, men som partikelliknande föremål. Det har vidare förutspåtts att det är möjligt att utnyttja Majoranafermioner för så kallade topologiska kvantberäkningar. Om det visar sig möjligt att både skapa Majoranafermioner och utnyttja dessa för sådana kvantberäkningar, så skulle detta kunna leda till en ny typ av dator som är många gånger mer kraftfull än dagens alla datorer tillsammans.

I de artiklar som ligger till grund för denna avhandling har vi studerat topologiska supraledare. Detta har till stor del gjorts med hjälp av numeriska beräkningar, vilka tillåter att mer komplicerade och verklighetstrogna modeller av topologiska supraledare kan studeras än vad som är möjligt med rent analytiska metoder. Vi undersöker bl.a. stabiliteten hos de topologiska faserna, hur olika parametrar påverkar existensen av Majoranafermioner, och hur man experimentellt skulle kunna identifiera dessa. I denna avhandling presenteras bakgrundsinformation som hjälper till att sätta dessa artiklar i perspektiv, ytterligare detaljer, och en sammanfattning av artiklarnas innehåll. I kapitel 2 introduceras viktiga matematiska begrepp inom differentialgeometri, topologi, och fiberknippen. Detta används sedan för att introducera topologisk bandteori i kapitel 3. En introduktion till supraledning ges i kapitel 4, vilken i kombination med kapitel 3 förbereder för en presentation av topologiska supraledare i kapitel 5. I kapitel 5 ges också en introduktion till den model som studerats i artiklarna. Kapitel 6 till 8 beskriver metodutveckling som skett parallellt med de mer modelspecifika beräkningarna, medan en sammanfattning av resultaten i artiklarna presenteras i kapitel 9.

12. Acknowledgments

I would like to thank Annica Black-Schaffer for giving me the opportunity to pursue a PhD and for excellent supervision. The freedom provided, in combination with frequent follow ups and almost instantaneous feedback is very appreciated. On the same note, I am thankful to Henrik Johannesson who, as a supervisor for my bachelor and master thesis, introduced me to the more theoretical aspects of condensed matter physics. In particular as this was at a time when material physics were among the branches of physics that was of least interest to me. I also appreciate comments on this thesis from Lucia Komendova, Mahdi Mashkoori, Jorge Cayao, Oladunjoye Awoga, Johann Schmidt, Jonas Fransson, and Annica Black-Schaffer, which have been very helpful for improving the readability, and Anna Sinelnikova for useful feedback about the front cover.

The material theory group at Uppsala University is a truly amazing place to be. I am therefore grateful to the senior people, who during many years in the past and present have worked hard to created this environment, which I during the past few years have been able to benefit immensely from. In the same way I am thankful to all people in the group, who make the everyday experience here very good. Thanks to Huiran, Manuel, Erna, Ritwik, Samara, Lucia, Henning, Johann, Tomas, Oscar, Henrik, Peter, Lisa, Annica, Anders, Jonas, Susanne, Juan, Fariborz, Johan, Adrien, Dushko, Saurabh, Francesco, and Anna for many hours of discussion in office, the lunch room, and elsewhere. Also, particular thanks to Alexander, Sergey, Mahdi, Oladunjoye, Leyla, Inka, Vancho, Sara, and Sofie for smooth collaborations with research and administration related tasks.

Finally, I would like to thank my family for their support. Many important ideas have developed during my stays back home, and in particular the summer and winter vaccations have been essential for the development of Paper IV, V, and VI. In fact, a special scientific thank goes to my mother for repeatedly helping with setting up port forward on the home router during the first half year of my PhD studies, without which Fig. 5 in Paper I would not have been possible. I also very much appreciate the interest my brother Benjamin recently have shown in math, physics, and electronics. Our regular discussions on these topics during the past year has meant a lot.

The research has been funded by the Swedish research council (VR), and SNIC has through Uppsala Multidisciplinary Center for Advanced Computational Science (Uppmax) provided computational resources.

References

[1] M. Epple, Arch. Hist. Exact Sci. **52**, 297 (1998).
[2] D. J. Thouless, *Topological Quantum Numbers In Nonrelativistic Physics* (World Scientific Publishing Co., 1997).
[3] N. D. Mermin, Rev. Mod. Phys. **51**, 591 (1979).
[4] A. Vilenkin and E. P. S. Shellard, *Cosmic strings and other topological defects* (1994).
[5] G. E. Volovik, *The Universe In A Helium Droplet* (Clarendon Press, 2003).
[6] T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).
[7] A. Altland and B. Simons, *Condensed matter field theory*, 3rd ed. (Cambridge University Press, 2011).
[8] K. v. Klitzing, Phys. Rev. Lett. **45**, 494 (1980).
[9] R. B. Laughlin, Phys. Rev. B **23**, 23 (1981).
[10] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. d. Nijs, Phys. Rev. Lett. **49**, 405 (1982).
[11] A. T. et al., Nature Nanotechnology **5**, 186 (2010).
[12] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[13] B. A. Bernevig and S.-C. Zhang, Phys. Re. Lett. **96**, 106802 (2006).
[14] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314** (2006).
[15] M. K. et al., Science **318**, 766 (2007).
[16] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature **452**, 970 (2008).
[17] Y. X. et al., Nature Physics **5**, 398 (2009).
[18] H. Z. et al., Nature Physics **5**, 438 (2009).
[19] Y. L. C. et al., Science **325**, 178 (2009).
[20] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
[21] X.-L. Qi and S.-C. Zhang, Mod. Rev. Phys. **83**, 1057 (2011).
[22] B. A. Bernevig and T. L. Hughes, *Topological insulators and topological superconductors* (Princeton University Press, 2013).
[23] L. Fu and C. L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).
[24] S. Fujimoto, Phys. Rev. B **77**, 220501 (2008).
[25] J. D. Sau, R. M. Lutchyn, S. Tewari, and S. D. Sarma, Phys. Rev. Lett. **104**, 040502 (2010).
[26] J. Alicea, Phys. Rev. B **81**, 125318 (2010).
[27] R. M. Lutchyn, J. D. Sau, and S. D. Sarma, Phys. Rev. Lett. **105**, 077001 (2010).
[28] Y. Oreg, G. Refael, and F. von Oppen, Phys. Rev. Lett. **105**, 177002 (2010).
[29] M. Sato, Y. Takahashi, and S. Fujimoto, Phys. Rev. Lett. **103**, 020401 (2009).
[30] M. Sato, Y. Takahashi, and S. Fujimoto, Phys. Rev. B **82**, 134521 (2010).
[31] V. Mourik, K. Zuo, S. M. Frolov, S. R. Plissard, E. P. A. M. Bakkers, and L. P. Kouwenhoven, Science **336**, 1003 (2012).

[32] E. Majorana, Nuevo Cimento **5**, 171 (1937).
[33] F. Wilczek, Nature Physics **5**, 614 (2009).
[34] A. Y. Kitaev, Physics Uspekhi **44**, 131 (2001).
[35] S. Nadj-Perge, I. K. Drozdov, J. Li, H. Chen, S. Jeon, J. Seo, A. H. MacDonald, B. A. Bernevig, and A. Yazdani, Science **346**, 602 (2014).
[36] R. Pawlak, M. Kisiel, J. Klinovaja, T. Meier, S. Kawai, T. Glatzel, D. Loss, and E. Meyer, arXiv:1505.06078 (2015).
[37] M. Ruby, F. Pientka, Y. Peng, F. von Oppen, B. W. Heinrich, and K. J. Franke, Phys. Rev. Lett. **115**, 197204 (2015).
[38] L. Landau, Ukr. J. Phys. **53**, 25 (2008).
[39] S. Weinberg, *Gravitation and cosmology* (John Wiley & Sons, Inc., 1972).
[40] I. Madsen and I. Tornehave, *From calculus to cohomology* (Cambridge University Press, 1997).
[41] M. Spivak, *Calculus on Manifolds* (Addison-Wesley Publishing Company, 1965).
[42] M. Nakahara, *Geometry, Topology and Physics*, 2nd ed. (Taylor & Francis, 2003).
[43] A. Pressley, *Elementary Differential Geometry*, 8th ed. (Springer, 2006).
[44] V. Pati, Resonance **1**, 37 (1996).
[45] M. A. Armstrong, *Basic Topology* (Springer-Verlag New York Inc., 1983).
[46] W. Rindler, *Relativity: Special, General, and Cosmological*, 2nd ed. (Oxfor University Press, 2006).
[47] G. K. Francis and J. R. Weeks, The American Mathematical Monthly **106**, 393 (1999).
[48] M. V. Berry, Proc. R. Soc. Lond. A **392**, 45 (1984).
[49] C. Kittel, *Introduction to SolidStatePhysics*, 8th ed. (John Wiley & Sons, Inc., 2005).
[50] N. W. Ashcroft and N. D. Mermin, *Solid state physics* (Harcourt College Publishing, 1976).
[51] Y. A. Bychkov and E. I. Rashba, JETP Letters **39**, 78 (1984).
[52] J. D. Bjorken and S. D. Drell, *Relativistic quantum mechanics*, customized ed. (McGraw-Hill, 1998).
[53] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Westview Press, 1995).
[54] T. O. Wehling, A. M. Black-Schaffer, and A. V. Balatsky, arxiv.1405.5774 (2014).
[55] C. L. Kane and E. J. Mele, Phys. Rev. Lett **95**, 146802 (2005).
[56] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
[57] J. R. Schrieffer, *Theory of Superconductivity* (Westview Press, 1999).
[58] P. G. de Gennes, *Superconductivity of Metals and Alloys* (Westview Press, 1999).
[59] M. Tinkham, *Introduction to superconductivity*, 2nd ed. (Dover, 2004).
[60] H. K. Onnes, Comm. Phys. Lab. Univ. Leiden, Nos 119, 120, 122 (1911).
[61] M. Sigrist and K. Ueda, Rev. Mod. Phys. **63**, 239 (1991).
[62] L. N. Cooper, Phys. Rev. **104**, 1189 (1956).
[63] N. N. Bogoliubov, Soviet Physics JETP **34**.
[64] J. G. Valatin, Il Nuevo Cimento **VII**.

[65] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of quantum field theory in statistical physics* (Pergamon Press Ltd., 1965).
[66] A. A. Abrikosov, J. Phys. Chem. Solids. **2**, 199 (1957).
[67] G. B. Arfken and H. J. Weber, *Mathematical Methods For Physicisists*, 6th ed. (Elsevier, 2005).
[68] A. Chrestin, T. Matsuyama, and U. Merkt, Phys. Rev. B **55**, 8457 (1997).
[69] E. Majorana, L. Maiani, and G. F. Bassani, *Ettore Majorana Scientific Papers* (Springer, 2006).
[70] G. E. Volovik, JETP Letters **70**, 609 (1999).
[71] N. Read and D. Green, Phys. Rev. B **61**, 10267 (2000).
[72] J. Alicea, Rep. Prog. Phys. **75**, 1 (2012).
[73] S. M. Sze and K. K. Ng, *Physics of semiconductor devices* (John Wiley & Sons, Inc., 2007).
[74] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. D. Sarma, Rev. Mod. Phys. **80**, 1083 (2008).
[75] A. B. et al., Phys. Rev. B. **52**, 52 (1995).
[76] I. M. Georgescu, S. Ashhab, and F. Nori, Rev. Mod. Phys. **86**, 152 (2014).
[77] L. Mao and C. Zhang, Phys. Rev. B **82**, 174506 (2010).
[78] D. A. Ivanov, Phys. Rev. Lett. **86**, 268 (2001).
[79] J. M. Leinaas and J. Myrheim, Il Nuovo Cimento B **37**, 1 (1977).
[80] J. Alicea, Y. Oreg, G. Refael, F. von Oppen, and M. P. A. Fisher, Nat. Phys. **7**, 412 (2011).
[81] J. P. Boyd, *Chebyshev and Fourier Spectral Methods* (Dover Publicaitons, Inc., 2000).
[82] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, Rev. Mod. Phys. **78**, 275 (2006).
[83] L. Covaci, F. M. Peeters, and M. Berciu, Phys. Rev. Lett. **105**, 1 (2010).
[84] G. D. Mahan, *Many-Particle Physics* (Plenum Press, 1990).
[85] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley & Sons, Inc., 1999).
[86] T. Daino, M. Ichioka, T. Mizushima, and Y. Tanaka, Phys. Rev. B **86**, 064512 (2012).
[87] H. Ebisu, K. Yada, H. Kasai, and Y. Tanaka, Phys. Rev. B **91**, 054518 (2015).
[88] Y. Asano and Y. Tanaka, Phys. Rev. B **87**, 104513 (2013).
[89] T. Tanaka, M. Sato, and N. Nagaosa, J. Phys. Soc. Jpn. **81**, 011013 (2012).
[90] H. P. Dahal, E. Abrahams, D. Mozyrsky, Y. Tanaka, and A. V. Balatsky, New J. Phys. **11**, 065005 (2009).
[91] A. M. Black-Schaffer and A. V. Balatsky, Phys. Rev. B **86**, 144506 (2012).
[92] X.-L. Qi, Y.-S. Wu, and S.-C. Zhang, Phys. Rev. B **74**, 085308 (2006).
[93] J. O. et al., *Magnetism: A Synchrotron Radiation Approach* (Springer, 2006).
[94] C. Caroli, P. G. de Gennes, and J. Matricon, Phys. Lett. **9**, 307 (1964).
[95] G. E. Volovik, J. Phys. C: Solid State Phys. **21**, L215 (1988).
[96] M. M. Salomaa and G. E. Volovik, J. Phys.: Condens. Matter **1**, 277 (1989).
[97] J. A. Sauls and M. Eshring, New J. Phys. **11**, 075008 (2009).
[98] M. Fogelström, Phys. Rev. B **84**, 064530 (2011).
[99] A. M. Black-Schaffer, Phys. Rev. B **88**, 104506 (2013).
[100] L. Yu, Acta Phys. Sin. **21**, 78 (1965).

[101] H. Shiba, Prog. Theor. Phys. **40**, 435 (1968).
[102] A. I. Rusinov, JETP Lett. **9**, 85 (1969).
[103] A. V. Balatsky, I. Vekhter, and J.-X. Zhu, Rev. Mod. Phys. **78**, 373 (2006).
[104] M. I. Salkola, A. V. Balatsky, and J. R. Schrieffer, Phys. Rev. B **55**, 12648 (1997).
[105] M. E. Flatté and J. M. Byers, Phys. Rev. Lett. **78**, 3761 (1997).
[106] T. Meng, J. Klinovaja, S. Hoffman, P. Simon, and D. Loss, Phys. Rev. B **92**, 064503 (2015).

Appendix A.
Chebyshev polynomial

Proof that $T_m(x)$ is a polynomial of degree m

Letting $T_m(x)$ be defined as in Eq. (7.5), and assuming $m > 1$, we have

$$\begin{aligned}
T_m(x) =& \text{Re}\left(e^{i(m-1)\theta}e^{i\theta}\right)\\
=& \cos((m-1)\theta)\cos(\theta) - \sin((m-1)\theta)\sin(\theta)\\
=& xT_{m-1}(x) - \text{Im}\left(e^{i(m-2)\theta}e^{i\theta}\right)\sin(\theta)\\
=& xT_{m-1}(x) - (\cos((m-2)\theta)\sin(\theta) + \sin((m-2)\theta)\cos(\theta))\sin(\theta)\\
=& xT_{m-1}(x) - \cos((m-2)\theta)\sin^2(\theta) - (\sin((m-2)\theta)\sin(\theta))\cos(\theta)\\
=& xT_{m-1}(x) - \cos((m-2)\theta)\\
&+ (\cos((m-2)\theta))\cos(\theta) - \sin((m-2)\theta)\sin(\theta))\cos(\theta)\\
=& xT_{m-1}(x) - T_{m-2}(x) + \text{Re}\left(e^{i(m-1)\theta}\right)x\\
=& 2xT_{m-1}(x) - T_{m-2}(x). \qquad (12.1)
\end{aligned}$$

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1441

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-305212

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016